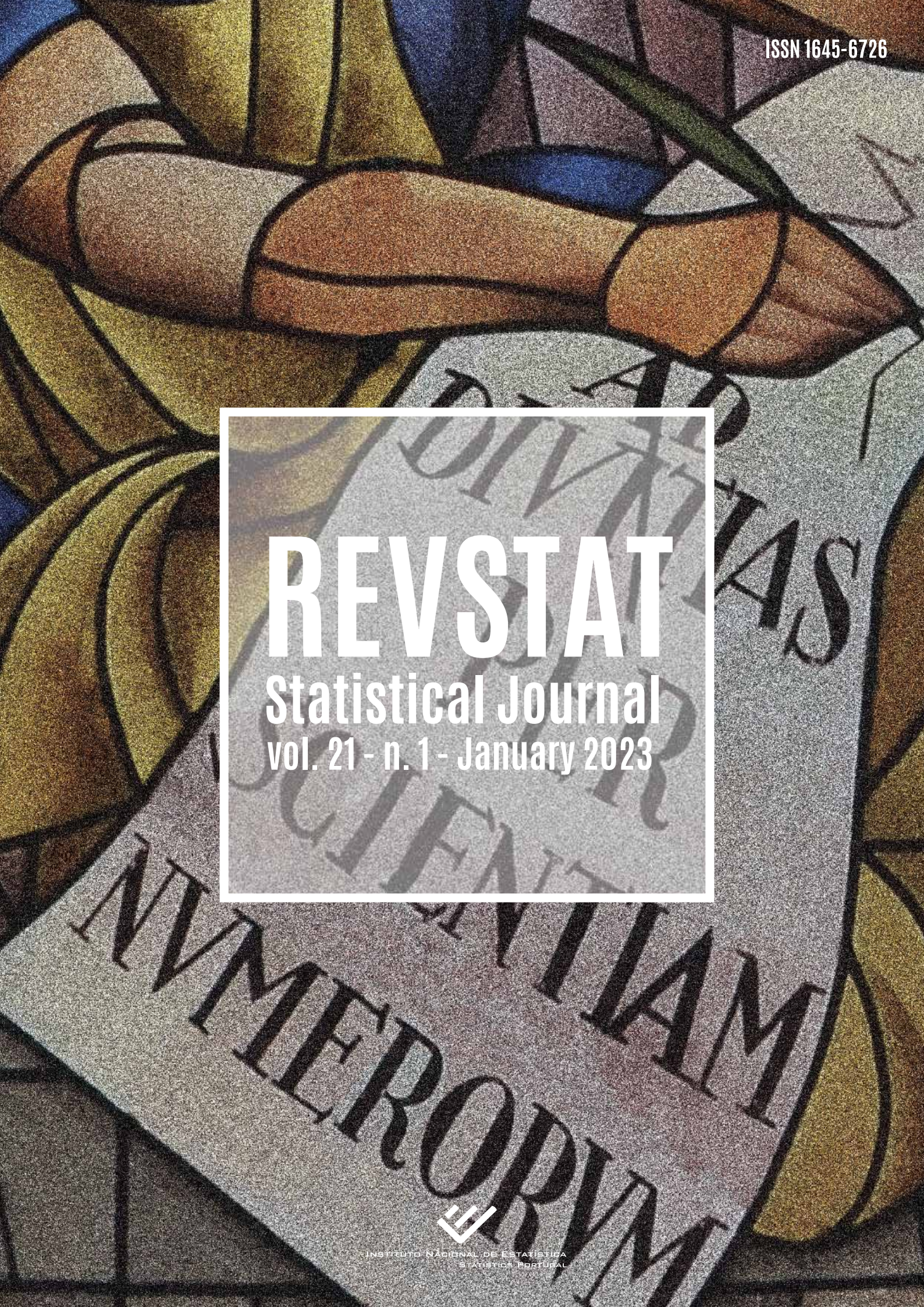
ISSN 1645-6726



REVSTAT

Statistical Journal

vol. 21 - n. 1 - January 2023

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT

Statistical Journal

vol. 21 - n. 1 - January 2023

REVSTAT-Statistical Journal, vol.21, n. 1 (January 2023)

vol.1, 2003- . - Lisbon : Statistics Portugal, 2003- .

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

Editorial Board (2019-2023)

Publisher – *Statistics Portugal*

Layout-Graphic Design – *Carlos Perpétuo* | **Cover Design*** – *Helena Nogueira*

Edition - 130 copies | **Legal Deposit Registration** - 191915/03 | **Price** [VAT included] - € 9,00

**image*: stain glass window by Abel Manta (1888-1982)

Editorial

In memoriam - Carlos Daniel Paulino [1951-2023]

On the 18th of April 2023, the statistical community in Portugal and abroad was chocked with the sad news that **Carlos Daniel Paulino** had passed away.

Although **Daniel** had a degree in Chemical Engineering, in his heart he had been always a mathematician. As such, he initiated his academic career in 1975 as an assistant in the Department of Mathematics at Instituto Superior Técnico (IST) in Lisbon, where he was attracted to the Statistical Science. In 1984 he completed the MSc. in Statistics and Operational Research, Faculty of Sciences of the University of Lisbon, with a dissertation on Bayesian Statistics, that later became his main focus of research. He took his Ph.D. degree in Statistics at the Institute of Mathematics and Statistics of the University of São Paulo in 1989. After returning to Portugal, he resumed his position in the Department of Mathematics at IST as an Assistant Professor and later in 1995 as an Associate Professor in Statistics, with habilitation in Statistics and Operational Research since 2005. In IST he lectured many subjects, both at postgraduate and undergraduate level. He retired from the IST in 2014, but he continued very active in research, as a Senior Researcher at the Centre of Statistics and its Applications (CEAUL), University of Lisbon.

Besides Bayesian Statistics, **Carlos Daniel Paulino** main research interests were on Statistical Inference, Categorical Data Analysis, Incomplete and Missing Data, Survival Analysis and Statistical Applications to Biomedical Sciences, subjects that he lectured, not only at the postgraduate level at IST, but also at other institutions as advanced courses, both in Portugal and abroad. He co-authored three books on Bayesian Statistics, a book on Categorical Data Analysis and a book of exercises in Probability and Statistics. He collaborated with researchers from Portugal, Brazil, United States, Mexico, Spain, Belgium and England. His last work, jointly with Julio Singer and their student Eliardo Costa, was recently published (2023) in Statistical Methods and Applications (DOI: 10.1007/s10260-022-00639-0). Although already ill, **Daniel** worked hard on the article, making a point of having it accepted for publication, before his very good friend, Julio Singer, who was very ill, would know about the publication before passing away, which actually happened in May 2022 (`https://revstat.ine.pt/index.php/REVSTAT/article/view/476`).

Daniel was a very enthusiastic communicator, a very dedicated teacher, a very generous supervisor, a very exigent researcher, a true scholar, and a staunch defender of the Portuguese language, as a scientific language and not only. Within the Portuguese Statistical Society (SPE),

he was one of the main contributors and promotor of the Statistics Glossary English-Portuguese (`https://www.spestatistica.pt/en/glossary`), which became a common project with the Brazilian Statistics Association (ABE). He was President of SPE from 2013-2015, being a promoter of Statistics activities for young people and driving the creation of the SPE Biometrics Section, which in turn promoted several scientific events together with the Galician Society for the Promotion of Statistics and Operational Research (SGAPEIO). He also served, with strong dedication and rigor, as an Associate Editor of REVSTAT – Statistical Journal from 2014-.

And finally making ours the words of Peter Müller '**Daniel** was one of our brightest heads, and a true academic. We'll miss him!'

May 24, 2023

Maria Antónia Amaral Turkman (*Former Co-Editor*)
Giovani L. Silva (*Current Co-Editor*)

INDEX

Abstracted/indexed in: *Journal Citation Reports™ – JCR (Clarivate); DOAJ – Directory of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; SCImago Journal & Country Rank; Scopus®.*

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 1–20
https://doi.org/10.57805/revstat.v21i1.396

A New Bivariate Birnbaum–Saunders Type Distribution Based on the Skew Generalized Normal Model

Authors: Barry C. Arnold
– Statistics Department, University of California Riverside, Riverside, USA
barry.arnold@ucr.edu

Diego I. Gallardo
– Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile
diego.gallardo@uda.cl

Héctor W. Gómez ✉
– Departamento de Estadística y Ciencias de Datos, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta, Chile
hector.gomez@uantof.cl

Received: April 2019 Revised: May 2021 Accepted: June 2021

Abstract:

- It is well known that it is possible to represent a Birnbaum–Saunders variable as a relatively simple (and invertible) function of a standard normal random variable. Marginal transformations of this kind are applied in this paper to a bivariate distribution with generalized skew-normal conditionals (and normal marginals), to obtain a new bivariate Birnbaum–Saunders distribution. Parameter estimation for this model is implemented using an EM algorithm. A simulation study sheds light on the performance of the estimation strategy. Data from a cancer risk study is used to illustrate use of the model. For this data set, the new model exhibits better performance than does a competing skew-normal based model already discussed in the literature. Possible multivariate extensions of the new model are outlined.

Keywords:

- *Birnbaum–Saunders distribution; bivariate distribution; conditional specifications; EM algorithm.*

AMS Subject Classification:

- 62E10, 62F10.

✉ Corresponding author.

1. INTRODUCTION

While there are many univariate models available for analysis of survival data, the same cannot be said for cases involving bivariate or, even more challenging, multivariate cases. In the univariate case there has been a flurry of recent activity focused on the Birnbaum–Saunders (BS) distribution (see, Birnbaum and Saunders [14]). A particularly attractive feature of the BS model is its representation as a monotone transformation of a standard normal variable.

Analogous distributions, which can be called distributions of the BS type, can be constructed by assuming that the normal random variable which is transformed to obtain a BS distributed variable, is replaced by a random variable with a different distribution. Recent papers dealing with the BS distribution and its close relatives include those of Balakrishnan and Kundu [13], Athayde [8], Bourguignon *et al.* [15], Arrué *et al.* [7], Carrasco *et al.* [16], Dasilva *et al.* [17] and Martínez-Flórez *et al.* [30]. See also the book by Leiva [27].

In particular we may consider replacing the normal component that is transformed to yield a BS variable by some skewed normal random variable. The traditional skew-normal (SN) distribution was introduced by Azzalini [9]. The skew-generalized-normal (SGN) distribution, introduced by Arellano-Valle *et al.* [2] (see also Arnold *et al.* [4]), includes an additional parameter. The SGN model can be viewed as a shape parameter mixture of SN distributions, where the shape parameter is endowed with a standard normal distribution. The model contains the SN model as particular case. The parameter space for the SGN distribution is $\{(\lambda, \theta) : -\infty < \lambda < \infty, \theta \geq 0\}$. As discussed in Arellano-Valle *et al.* [2], this model has identifiability problems which can be circumvented by restricting the parameter space, resulting in a distribution known as the skew-curved-normal (SCN) distribution (see, Gómez *et al.* [23]) . It is this SCN distribution that we propose to use instead of a standard normal distribution in order to arrive at a flexible extension of the BS model, which we will call a skew-curved-normal-BS (SCNBS) model.

While there has been much discussion of univariate variations on the BS theme, there is much less available for analyzing higher dimensional survival data. The present paper will make a contribution towards filling this gap. Our interest, then, is in the development of flexible bivariate and multivariate BS distributions. The bivariate case will be discussed in detail. Our goal is to seek models which have BS marginals and, in addition, will exhibit well behaved conditional structure. As will be seen, approaches involving conditional specification of joint distributions will prove to be fruitful. A convenient reference for discussion of conditionally specified models is Arnold *et al.* [6].

The paper will be organized as follows. Section 2 reviews the construction of the univariate BS distribution and its variants, and introduces the new bivariate SCNBS (BSCNBS) model and its main properties. Section 3 presents the inference for the BSCNBS model based on a classical approach and includes discussion of residuals for this model from both a marginal and a bivariate point of view. Section 4 includes a simulation study to assess the performance of the estimators obtained with the EM algorithm in finite samples and includes a real data application to illustrate the performance of the BSCNBS model. Multivariate extensions are not difficult to envision and are described briefly in Section 5. In Section 6 we present the main conclusions of the paper, together with discussion of related topics.

2. THE MODEL

In this Section, we introduce the BSCNBS model, where the conditional distributions are SCNBS and the marginal distribution are BS. Some properties of the model are discussed, as is a procedure to draw values from the model.

2.1. A background of related univariate distributions

A random variable T is said to have a BS(α, β) distribution if it can be represented in the form

$$T = \beta\left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2}\right)^2 + 1}\right]^2, \tag{2.1}$$

where $Z \sim$N$(0, 1)$, i.e., the standard normal distribution. The density function of such a random variable is given by

$$f_T(t) = \phi\left(\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right)\frac{t^{-3/2}(t+\beta)}{2\alpha\sqrt{\beta}}, \qquad t > 0, \tag{2.2}$$

where ϕ denotes the density function for the standard normal distribution. As mentioned in the introduction, we may replace Z by some skewed normal random variable.

The density function of a random variable with SGN distribution is given by

$$f_U(u) = 2\phi(u)\Phi\left(\frac{\lambda u}{\sqrt{1+\theta u^2}}\right), \qquad u \in \mathbb{R}, \tag{2.3}$$

where Φ denotes the cumulative distribution function for the standard normal model. Note that if $\theta = 0$, this simplifies to the form of the traditional SN model. The parameter space for the SGN distribution is $\{(\lambda, \theta) : -\infty < \lambda < \infty, \theta \geq 0\}$, which is reduced to the SCN model for $\theta = \lambda^2$.
In addition, the SCNBS model is obtained considering Z with SCN distribution in the transformation in equation (2.1). The associated density function for the SCNBS distribution is

$$f_T(t) = 2\phi(a)\Phi\left(\frac{\lambda a}{\sqrt{1+\lambda^2 a^2}}\right)A, \qquad t > 0,$$

where $a = \alpha^{-1}\left[(t/\beta)^{1/2} - (\beta/t)^{1/2}\right]$ and $A = t^{-3/2}(t+\beta)/(2\alpha\sqrt{\beta})$. We use the notation SGN(λ, θ) to refer to a random variable with this density function. Those distributions are very relevant to the construction of the our proposal.

2.2. A Bivariate SCNBS distribution

Before describing the proposed bivariate distribution, which has Birnbaum Saunders marginals and SCNBS conditional distributions, we review two bivariate BS distributions that

have been discussed in the literature. As will be seen, both the existing bivariate BS models and the one proposed in this paper are constructed by means of marginal transformations applied to bivariate densities with normal marginals and normal or skew-normal conditionals.

The first bivariate BS (BVBS) model was proposed by Kundu *et al.* [26]. They began by assuming that (Z_1, Z_2) has a classical bivariate normal distribution with standard normal marginals and correlation ρ. They then defined, as in (2.1)

$$T_i = \beta_i \left[\frac{\alpha_i Z_i}{2} + \sqrt{\left(\frac{\alpha_i Z_i}{2}\right)^2 + 1} \right]^2, \quad i = 1, 2. \tag{2.4}$$

It is evident that the bivariate random variable (Z_1, Z_2) so defined has BS marginal distributions and BS conditional distributions.

In order to enhance the flexibility of the model (2.4), Lemonte *et al.* [28] proposed a more general model using a parallel construction which utilizes a distribution with SN conditionals introduced by Arnold *et al.* [3]. This bivariate density is of the form

$$f(z_1, z_2) = 2\phi(z_1)\phi(z_2)\Phi(\lambda z_1 z_2), \qquad (z_1, z_2) \in \mathbb{R}^2. \tag{2.5}$$

Lemonte *et al.* [28] then apply the marginal transformation (2.4) to a bivariate random variable (Z_1, Z_2) with joint density of the form (2.5). Since the Arnold *et al.* [3] density (2.5) is readily verified to have standard normal marginals and SN conditionals, it follows that the Lemonte *et al.* [28] distribution will have BS marginals and skew-normal-Birnbaum–Saunders (SNBS) conditionals. The bivariate SNBS (BSNBS) density studied by Lemonte *et al.* [28] is

$$f_{T_1,T_2}(t_1, t_2) = 2\phi(a_1)\phi(a_2)\Phi(\lambda a_1 a_2) A_1 A_2, \qquad (t_1, t_2) \in \mathbb{R}^2_+,$$

where $a_j = a_j(\alpha_j, \beta_j) = \alpha_j^{-1}\left[\sqrt{t_j/\beta_j} - \sqrt{\beta_j/t_j}\right]$ and

$$A_j = A_j(\alpha_j, \beta_j) = t_j^{-3/2}(t_j + \beta_j)/(2\alpha_j\sqrt{\beta_j}), \quad j = 1, 2.$$

In addition to the density (2.5), Arnold *et al.* [4] proposed a more general two parameter model of the form

$$f(z_1, z_2) = 2\phi(z_1)\phi(z_2)\Phi\left(\frac{\lambda z_1 z_2}{\sqrt{1 + \theta z_1^2 z_2^2}}\right), \qquad (z_1, z_2) \in \mathbb{R}^2_+, \tag{2.6}$$

where $\lambda \in (-\infty, \infty)$ and $\theta \in [0, \infty)$. In this paper, we will consider the case $\theta = \lambda^2$. This distribution then has standard normal marginals and has generalized skew-normal conditional distributions. Specifically we have, if (Z_1, Z_2) has density (2.6) with $\theta = \lambda^2$ then

$$Z_1 | Z_2 = z_2 \sim \mathrm{SCN}(\lambda z_2) \tag{2.7}$$

and

$$Z_2 | Z_1 = z_1 \sim \mathrm{SCN}(\lambda z_1). \tag{2.8}$$

It is to this joint distribution that we apply the marginal transformations (2.4) to obtain a flexible bivariate distribution with BS marginals that will be the focus of the remainder of this paper. The resulting joint density is of the form

$$f_{T_1,T_2}(t_1, t_2) = 2\phi(a_1)\phi(a_2)\Phi\left(\frac{\lambda a_1 a_2}{\sqrt{1 + \lambda^2 a_1^2 a_2^2}}\right) A_1 A_2, \qquad (t_1, t_2) \in \mathbb{R}^2_+. \tag{2.9}$$

If a random variable (T_1, T_2) has its density of the form (2.9) then we write $(T_1, T_2) \sim \mathrm{BSCNBS}(\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda)$.

For the BSCNBS$(\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda)$ distribution, we have the following properties:

(**1**) $T_i \sim$BS(α_i, β_i), $\quad i = 1, 2$.

(**2**) $T_1 \,|\, T_2 = t_2 \sim$SCNBS$(\alpha_1, \beta_1, \lambda a_2)$ and $T_2 \,|\, T_1 = t_1 \sim$SCNBS$(\alpha_2, \beta_2, \lambda a_1)$.

(**3**) $(c_1 T_1, c_2 T_2) \sim$BSCNBS$(\alpha_1, \alpha_2, c_1\beta_1, c_2\beta_2, \lambda), c_i > 0,\ i = 1, 2$.

(**4**) $(c_1 T_1, c_2 T_2^{-1}) \sim$BSCNBS$(\alpha_1, \alpha_2, c_1\beta_1, c_2\beta_2^{-1}, -\lambda), c_i > 0,\ i = 1, 2$.

(**5**) $(c_1 T_1^{-1}, c_2 T_2) \sim$BSCNBS$(\alpha_1, \alpha_2, c_1\beta_1^{-1}, c_2\beta_2, -\lambda), c_i > 0,\ i = 1, 2$.

(**6**) $(c_1 T_1^{-1}, c_2 T_2^{-1}) \sim$BSCNBS$(\alpha_1, \alpha_2, c_1\beta_1^{-1}, c_2\beta_2^{-1}, \lambda), c_i > 0,\ i = 1, 2$.

(**7**) And, going back, if $Z_i = \alpha_i^{-1}\left[\sqrt{T_i/\beta_i} - \sqrt{\beta_i/T_i}\right]$, $i = 1, 2$, then $Z_1 \,|\, Z_2 = z_2 \sim$ SCN(λz_2) and $Z_2 \,|\, Z_1 = z_1 \sim$SCN(λz_1). Using Proposition 10 in Arellano-Valle *et al.* [2], the first conditional distribution is equivalent to $Z_1 \,|\, Z_2 = z_2, U = u \sim$SN$(u)$ and $U \sim$N$(\lambda z_2, \lambda^2 z_2^2)$. Similarly, $Z_2 \,|\, Z_1 = z_1, U = u \sim$SN$(u)$ and $U \sim$N$(\lambda z_1, \lambda^2 z_1^2)$.

Parts (1) and (2) are obtained directly from the definition of the distribution and the results given in (3) to (6). are obtained using appropriate transformations in the density given in (2.9). The representation given in (7) of the conditional distributions are useful for simulation purposes, as illustrated in the following Sub-Section. Figure 1 shows the contour levels and Figure 2 shows the density for BSCNBS model for some combinations of the parameters. Note that the contours exhibit different and a greater variety of shapes than the BSNBS model.

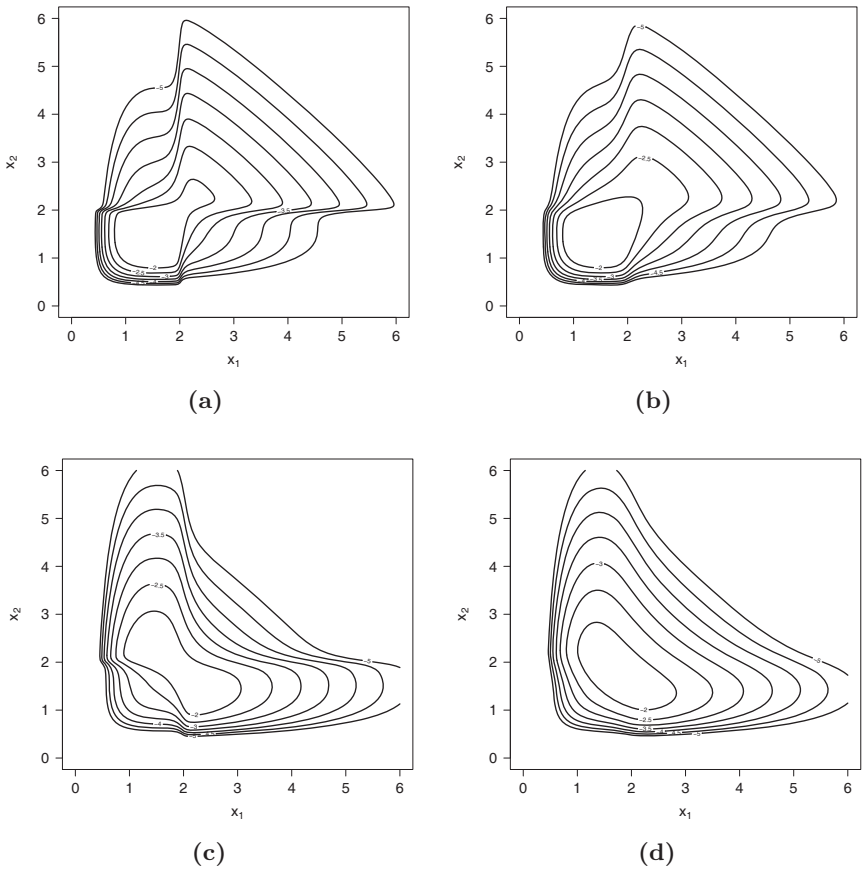


Figure 1: Contours levels for BSCNBS$(\alpha_1 = 0.5, \alpha_2 = 0.5, \beta_1 = 2, \beta_2 = 2, \lambda)$ distribution considering: (a) $\lambda = 7$; (b) $\lambda = 2.5$; (c) $\lambda = -3.5$ and; (d) $\lambda = -1$.

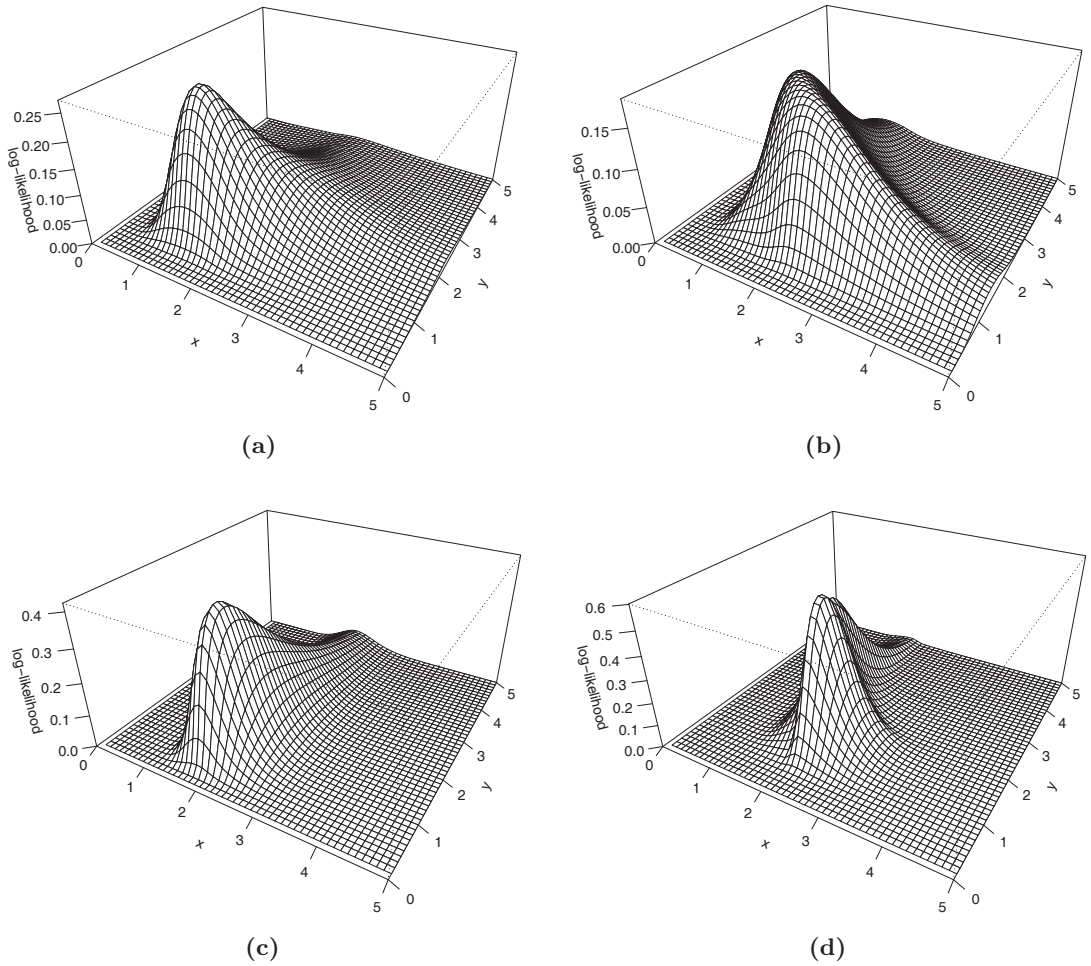


Figure 2: Density for BSCNBS$(\alpha_1, \alpha_2, \beta_1 = 2, \beta_2 = 2, \lambda)$ distribution considering: (a) $\alpha_1 = \alpha_2 = 0.5$ and $\lambda = 1$; (b) $\alpha_1 = \alpha_2 = 0.5$ and $\lambda = -1$; (c) $\alpha_1 = 0.2$, $\alpha_2 = 0.7$ and $\lambda = 0.5$; (d) $\alpha_1 = 0.2, \alpha_2 = 0.4$ and $\lambda = -2$.

The non-singularity of the Fisher information matrix (FIM) for $\lambda = 0$ is verified in the Appendix. This point is very important because $\lambda = 0$ represents the case where the model is reduced to two independent BS variates. Therefore, the non-singularity of the FIM allows to apply usual hypothesis test such as maximum likelihood ratio, score and Wald tests to decide between BSCNBS and independent BS variates.

2.3. Drawn values from BSCNBS distribution

The fact that this model has conditional and marginal distributions in closed form allows one to draw values from the distribution BSCNBS in a relatively simple way using the following Algorithm 1.

Algorithm 1 Simulate a value from the BSCNBS$(\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda)$ distribution.

1: Draw $Z_1 \sim$N$(0,1)$.
2: Draw $Z_2 \,|\, Z_1 = z_1 \sim$ SCN(λz_1).
  2.1: Draw $U \sim$ N$(\lambda z_1, \lambda^2 z_1^2)$.
  2.2: Draw $Z_2 \,|\, U = u \sim$ SN(u).
    2.2.1: Draw $V_1, V_2 \sim$ N$(0,1)$ (independent).
    2.2.2: Let $Z_2 = \left(\frac{u}{\sqrt{1+u^2}}\right)|V_1| + \left(\frac{1}{\sqrt{1+u^2}}\right)V_2$.
3: Make the usual BS-type transformation

$$T_j = \beta_j \left[\frac{\alpha_j Z_j}{2} + \sqrt{\left(\frac{\alpha_j Z_j}{2}\right)^2 + 1} \right]^2, \quad j = 1, 2.$$

Remark 2.1. This algorithm requires only obvious minor modification to produce a drawn value from the Lemonte *et al.* [28] distribution.

3. ESTIMATION

In this Section we consider the parameter estimation problem based on a classical approach. An EM algorithm is developed for this problem. Initial values for such procedures and two kind of residuals also are presented.

3.1. Estimation based on the EM algorithm

For the BSCNBS model, the log-likelihood function for $\boldsymbol{\psi} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda)$ in a random sample $\boldsymbol{t} = \boldsymbol{t}_1, \boldsymbol{t}_2, ..., \boldsymbol{t}_n$ (where $\boldsymbol{t}_i = (t_{i1}, t_{i2})$) is, up to a constant, given by

$$\ell(\boldsymbol{\psi}) = -\frac{n}{2}\sum_{j=1}^{2}\left[\frac{1}{\alpha_j^2}\left(\frac{\bar{s}_j}{\beta_j} + \bar{r}_j\beta_j - 2\right) + 2\log(\alpha_j) + \log(\beta_j)\right] + \sum_{j=1}^{2}\sum_{i=1}^{n}\log(\beta_j + t_{ij})$$
$$+ \sum_{i=1}^{n}\log\Phi\left(\frac{\lambda a_{i1}a_{i2}}{\sqrt{1 + (\lambda a_{i1}a_{i2})^2}}\right), \tag{3.1}$$

where

$$\bar{s}_j = \frac{1}{n}\sum_{i=1}^{n} t_{ji} \qquad \text{and} \qquad \bar{r}_j = \left(\frac{1}{n}\sum_{i=1}^{n} t_{ji}^{-1}\right)^{-1}, \qquad j = 1, 2.$$

and $a_{ij} = \alpha_j^{-1}\left[(t_{ij}/\beta_j)^{1/2} - (\beta_j/t_{ij})^{1/2}\right]$. The maximum likelihood (ML) estimation requires the maximization of eq. (3.1) in relation to $\boldsymbol{\psi}$. However, such a procedure can be difficult to implement because it involves maximization over a parameter space of dimension 5. For this reason, we discuss the development of an EM-type algorithm (Dempster *et al.* [18]) for this problem. This algorithm has been applied satisfactorily in BS models and their extension by Balakrishnan *et al.* ([10],[11],[12]), Pradhan and Kundu [34], Reyes *et al.* ([36], [37]), Romeiro *et al.* [38], among others.

A hierarchical representation of the BSCNBS model is given by

$$\begin{aligned} T_{1i} \,|\, T_{2i} = t_{2i}, U_i = u_i &\sim \text{SCNBS}(\alpha_1, \beta_1, u_i) \\ U_i \,|\, T_{2i} = t_{2i} &\sim \text{N}(\lambda a_{2i}, \lambda^2 a_{2i}^2) \\ T_{2i} &\sim \text{BS}(\alpha_2, \beta_2), \quad i = 1, ..., n. \end{aligned} \tag{3.2}$$

Let $\boldsymbol{t}$ and $\boldsymbol{u} = (u_1, ..., u_n)$ the observed values and the unobserved latent values, respectively. The complete data set then is $\mathbf{t}_c = (\mathbf{t}^T, \mathbf{u}^T)^T$. Using (3.2), the log-likelihood of the complete data set is given by

$$\begin{aligned} \ell_c(\boldsymbol{\psi} \,|\, \boldsymbol{t}_c) \propto & \sum_{j=1}^{2} \sum_{i=1}^{n} \left[-\frac{1}{\alpha_j^2} \left(\frac{t_{ji}}{\beta_j} + \frac{\beta_j}{t_{ji}} - 2 \right) + \log(t_{ji} + \beta_j) - \log(\alpha_j) - \frac{1}{2} \log(\beta_j) \right] \\ & - n \log \lambda + n \log \alpha_2 - \frac{1}{2} \sum_{i=1}^{n} \Big[\log \left(\frac{t_{1i}}{\beta_1} + \frac{\beta_1}{t_{1i}} - 2 \right) + \log \Phi(u_i a_{1i}) \\ & - \frac{1}{2\lambda^2 a_{2i}^2} (u_i^2 + \lambda^2 a_{2i}^2 - 2\lambda u_i a_{2i}) \Big]. \end{aligned}$$

Let $\widehat{u}_i^k = E(U_i^k \,|\, t_i, \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}})$, $k = 1, 2$. Note that

$$\begin{aligned} f(u_i \,|\, t_{1i}, t_{2i}, \boldsymbol{\psi}) &\propto f(t_{1i} \,|\, t_{2i}, u_i, \boldsymbol{\psi}) f(u_i \,|\, \boldsymbol{\psi}), \\ &\propto \phi \left(\frac{u_i - \lambda a_{2i}}{\lambda a_{2i}} \right) \Phi(u_i a_{1i}), \qquad u_i \in \mathbb{R}. \end{aligned}$$

Defining $C_{ki} = \int_{-\infty}^{\infty} u_i^k \phi\big((\lambda a_{2i})^{-1}(u_i - \lambda a_{2i})\big) \Phi(u_i a_{1i}) \mathrm{d}u_i$, we have $\widehat{u}_i^r = C_{ri}/C_{0i}$, $r = 1, 2$. Note that the existence of C_{ri} is guaranteed since

$$\int_{-\infty}^{\infty} u_i^r \phi \left(\frac{u_i - \lambda a_{2i}}{\lambda a_{2i}} \right) \Phi(u_i a_{1i}) \mathrm{d}u_i \leq \lambda a_{2i} \int_{-\infty}^{\infty} u_i^r \frac{1}{\lambda a_{2i}} \phi \left(\frac{u_i - \lambda a_{2i}}{\lambda a_{2i}} \right) \mathrm{d}u_i < \infty.$$

In this manner, the estimation process for this model, using the EM algorithm, may be described as follows in Algorithm 2.

The process is repeated iteratively until convergence is attained. For instance, we considered $\epsilon = 10^{-4}$.

Remark 3.1.

- **i)** The application of the ECM algorithm requires only uni-dimensional procedures, instead of the original problem which required a maximization of dimension 5.
- **ii)** The integrations involved in the E-step can be easily computed in the R software (R Core Team, [35]) with the `integrate` function.
- **iii)** The CM steps of the algorithm explicitly update λ, α_1 and α_2 and require the solution of a non-linear equation for β_1 and β_2. Such equations can be solved using the `uniroot` function in the R software.

Algorithm 2 Provide the ML estimates based on the EM algorithm for the BSCNBS distribution.

Set initial values $\boldsymbol{\psi}^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \lambda^{(0)})$
$k \leftarrow 0$
dif $\leftarrow 1$
while dif $> \epsilon$ **do**
  $i \leftarrow 1$
  while $i \leq n$ **do**
    (E-step) Compute the expected values for U_i and U_i^2
    $\widehat{u}_i^{(k+1)} = \frac{C_{1i}^{(k)}}{C_{0i}^{(k)}}$ and $\widehat{u}_i^2 = \frac{C_{2i}^{(k)}}{C_{0i}^{(k)}}$.
    $i \leftarrow i + 1$
  end while
  (CM-step I) Update λ

$$\widehat{\lambda}^{(k+1)} = \frac{\sum_{i=1}^{n} u_i^{(k+1)}}{\sum_{i=1}^{n} a_{2i}^{(k)}}.$$

  $j \leftarrow 1$
  while $j \leq 2$ **do**
    (CM-step II) Update α_j
    $\widehat{\alpha}_j^{2(k+1)} = \frac{S_j}{\widehat{\beta}_j^{(k)}} + \frac{\widehat{\beta}_j^{(k)}}{R_j} - 2$
    (CM-step III) Update β_j as the solution of the equation
    $\widehat{\beta}_j^{2(k+1)} - \widehat{\beta}_j^{(k+1)}\Big[K_j(\widehat{\beta}_j^{(k+1)}) + 2R_j\Big] + R_j\Big[K_j(\widehat{\beta}_j^{(k+1)}) + S_j\Big] = 0$
    where $K_j(x) = \big\{\frac{1}{n}\sum_{i=1}^{n}(x + t_{ji})\big\}^{-1}$.
    $j \leftarrow j + 1$
  end while
  $\boldsymbol{\psi}^{(k+1)} = (\alpha_1^{(k+1)}, \alpha_2^{(k+1)}, \beta_1^{(k+1)}, \beta_2^{(k+1)}, \lambda^{(k+1)})$.
  dif$= ||\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}^{(k)}||$, where $||x||$ denotes the Euclidean norm of the vector x.
  $k \leftarrow k + 1$
end while

3.2. Initial values of the algorithm

Since $T_j \sim$BS(α_j, β_j), we can use modified moment estimators of the BS distribution to estimate α_j and β_j, $j = 1, 2$ (see Ng *et al.* [33]). Thus, the initial values for those parameters are

$$\widehat{\alpha}_j^{(0)} = \sqrt{2\left(\sqrt{\frac{\overline{s}_j}{\overline{r}_j}} - 1\right)} \quad \text{and} \quad \widehat{\beta}_j^{(0)} = \sqrt{\overline{s}_j \overline{r}_j}, \qquad j = 1, 2. \tag{3.3}$$

With those values, we can construct a profile version of (3.1) for λ and choose the value of λ that maximizes that function.

3.3. Residuals for the BSCNBS model

In order to check the goodness of fit of the BSCNBS model, we evaluate the marginal quantile residuals (MQR; Dunn and Smyth, [19]) and the bivariate quantile residuals (BQR; Kalliovirta, [24]). Such theoretical residuals are given by

$$r_{ij}^{\mathrm{MQR}} = a_{ij} \quad \text{and} \quad r_i^{\mathrm{BQR}} = \Phi^{-1}(\nu_i(1-\log\nu_i)),$$

respectively, for $i = 1, ..., n, j = 1, 2$, where

$$\nu_i = \Phi(a_{i1}) \int_{-\infty}^{a_{i2}} 2\phi(u)\Phi\left(\frac{\lambda a_{1i} u}{\sqrt{1+\lambda^2 a_{1i}^2 u^2}}\right) \mathrm{d}u, \quad i = 1, ..., n,$$

where $a_{ij} = \alpha_j^{-1}\left[(t_{ij}/\beta_j)^{1/2} - (\beta_j/t_{ij})^{1/2}\right]$, $i = 1, ..., n$, $j = 1, 2$. The observed MQR and BQR (say $\widehat{r}_{i1}^{\mathrm{MQR}}$, $\widehat{r}_{i2}^{\mathrm{MQR}}$ and $\widehat{r}_i^{\mathrm{BQR}}$) are the theoretical MQR and BQR, respectively, evaluated as functions of the estimated parameters.
If the BS model is correctly specified for the j-th variable, then $\widehat{r}_{1j}^{\mathrm{MQR}}, ..., \widehat{r}_{nj}^{\mathrm{MQR}}$ has a $\mathrm{N}(0,1)$ distribution.
In a similar way, if the BSCNBS model is correctly specified for the two variables (jointly), then $r_1^{\mathrm{BQR}}, ..., r_n^{\mathrm{BQR}}$ has a $\mathrm{N}(0,1)$ distribution. Such hypothesis can be tested considering, for instance, the Kolomogorov–Smirnov (KS; Kolmogorov, [25]) test.

4. NUMERICAL RESULTS

In this Section we present details computational aspects used for this work. We also present a simulation study to assess the performance of the ML estimators obtained by the ECM algorithm discussed previously and a real data illustration in order to show the performance of the BSCNBS model. For the sake of comparison, we also consider the BSNBS model of Lemonte *et al.* [28] and the BVBS model of Kundu *et al.* [26].

4.1. Computational aspects

All the programs used in this work were run in R Core Team [35] in a computer with processor Intel(R) Core(TM) i7-7700HQ CPU 2.8GHz with 16 GB of RAM memory. The used packages for the development were the `VGAM` package [40] which includes some functions related to the BS model, the `mvtnorm` package [22] which includes some functions related to the multivariate normal model, the `goftest` package [20] including some functions related to goodness-of-fit tests and the `DAAG` package [29] which include the data used in the application presented in subsection 4.3. Codes for the application are included as supplementary material.

4.2. Simulation study

In this Section we report on a small simulation study with the objective of verifying that the EM-based estimation procedure is capable of recovering, approximately, the parameter values used to simulate data sets from the model (2.9). To simulate the data sets, we use the procedure described in Section 2.1. Then we use as the initial values, those discussed in Section 3.1 together with the EM algorithm outlined in Section 3. In particular we consider the parameter values $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$ in all cases while λ ranges over the set $\{-5, -2, -1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1, 2, 5\}$. In addition, we consider three sample sizes: $n = 100$, $n = 250$ and $n = 500$. In each case we make $1,000$ replications and calculate the mean absolute bias (AB), the mean of the standard errors (SE_1), the standard deviation of the estimated parameters (SE_2), and the coverage proportion (CP) of the nominal 95% intervals for the parameters. The results are presented in Tables 1 and 2. In these Tables we see that the biases of the estimates of $\alpha_1, \alpha_2, \beta_1$ and β_2 are negligible in all cases considered. However, the bias of the estimates of λ can be considerable in cases in which the true value of λ is far from 0. Although, as expected, the biases decrease as sample size increases.

Table 1: Simulation study for the BSCNBS model.

Case	Parameter	$n = 100$				$n = 250$				$n = 500$			
		AB	SE_1	SE_2	CP	AB	SE_1	SE_2	CP	AB	SE_1	SE_2	CP
$\lambda = -5.00$	α_1	−0.007	0.070	0.073	0.940	−0.004	0.045	0.044	0.952	−0.001	0.032	0.031	0.956
	α_2	−0.005	0.071	0.072	0.936	−0.004	0.045	0.047	0.935	−0.001	0.032	0.033	0.941
	β_1	0.001	0.067	0.072	0.908	−0.001	0.042	0.043	0.943	0.000	0.030	0.031	0.939
	β_2	0.005	0.069	0.072	0.921	0.002	0.042	0.044	0.924	0.001	0.029	0.030	0.931
	λ	−1.758	5.891	5.391	0.863	−1.075	3.030	3.371	0.935	−0.637	1.847	2.510	0.943
$\lambda = -2.00$	α_1	−0.006	0.070	0.069	0.939	−0.004	0.045	0.044	0.959	−0.001	0.032	0.032	0.939
	α_2	−0.006	0.070	0.070	0.939	−0.003	0.045	0.044	0.947	−0.002	0.032	0.032	0.949
	β_1	−0.001	0.079	0.079	0.943	0.002	0.050	0.051	0.945	0.000	0.035	0.035	0.944
	β_2	0.004	0.079	0.082	0.925	0.003	0.050	0.052	0.936	0.001	0.035	0.036	0.938
	λ	−0.725	2.052	2.679	0.904	−0.263	0.875	1.071	0.931	−0.128	0.543	0.572	0.956
$\lambda = -1.00$	α_1	−0.009	0.070	0.070	0.930	−0.005	0.045	0.044	0.942	−0.003	0.032	0.032	0.947
	α_2	−0.010	0.070	0.070	0.941	−0.001	0.045	0.045	0.943	−0.001	0.032	0.032	0.940
	β_1	0.008	0.084	0.087	0.933	0.003	0.053	0.053	0.952	0.000	0.038	0.037	0.947
	β_2	0.004	0.084	0.087	0.937	0.000	0.053	0.055	0.937	0.001	0.038	0.037	0.954
	λ	−0.324	0.842	1.237	0.906	−0.074	0.377	0.409	0.927	−0.036	0.251	0.260	0.941
$\lambda = -0.75$	α_1	−0.011	0.070	0.069	0.935	−0.004	0.045	0.046	0.949	−0.001	0.032	0.029	0.957
	α_2	−0.009	0.070	0.074	0.918	−0.003	0.045	0.044	0.944	−0.002	0.032	0.032	0.939
	β_1	0.008	0.086	0.089	0.934	0.001	0.054	0.053	0.958	−0.002	0.038	0.039	0.941
	β_2	−0.003	0.085	0.088	0.937	0.001	0.054	0.057	0.938	0.001	0.038	0.038	0.954
	λ	−0.175	0.559	0.773	0.913	−0.046	0.283	0.304	0.941	−0.025	0.191	0.198	0.942
$\lambda = -0.50$	α_1	−0.008	0.070	0.074	0.930	−0.003	0.045	0.046	0.948	−0.003	0.032	0.031	0.952
	α_2	−0.002	0.071	0.070	0.947	−0.003	0.045	0.045	0.949	−0.002	0.032	0.031	0.952
	β_1	0.006	0.087	0.087	0.947	0.001	0.055	0.056	0.946	0.002	0.039	0.040	0.944
	β_2	0.004	0.087	0.086	0.947	−0.002	0.055	0.054	0.945	−0.001	0.039	0.039	0.952
	λ	−0.150	0.410	0.565	0.931	−0.039	0.203	0.220	0.947	−0.020	0.137	0.143	0.945
$\lambda = -0.25$	α_1	−0.010	0.070	0.070	0.942	−0.003	0.045	0.044	0.935	−0.001	0.032	0.031	0.953
	α_2	−0.007	0.070	0.070	0.951	−0.002	0.045	0.045	0.945	−0.002	0.032	0.032	0.950
	β_1	0.002	0.087	0.089	0.947	0.003	0.056	0.055	0.955	0.001	0.039	0.041	0.939
	β_2	0.007	0.088	0.087	0.946	−0.001	0.055	0.056	0.942	0.003	0.039	0.039	0.944
	λ	−0.054	0.229	0.261	0.945	−0.023	0.130	0.135	0.955	−0.006	0.087	0.089	0.950
$\lambda = 0.00$	α_1	−0.009	0.070	0.069	0.949	−0.004	0.045	0.044	0.951	0.000	0.032	0.031	0.960
	α_2	−0.010	0.070	0.070	0.938	−0.001	0.045	0.044	0.947	−0.002	0.032	0.031	0.952
	β_1	0.001	0.087	0.089	0.943	0.000	0.056	0.054	0.951	−0.001	0.039	0.041	0.933
	β_2	0.004	0.088	0.088	0.939	0.005	0.056	0.056	0.941	0.001	0.039	0.039	0.938
	λ	0.006	0.156	0.176	0.996	0.001	0.086	0.086	0.990	0.002	0.059	0.057	0.978

Note that the values of SE_1 y SE_2 are very similar for $\alpha_1, \alpha_2, \beta_1$ and β_2 in all cases considered, which suggests that the standard errors of the estimates are themselves well estimated. However, for estimates of λ in most cases we have $SE_2 > SE_1$, suggesting that the standard errors of the λ estimates are underestimated, especially, once again, when the true value of λ is far from 0. We note that the coverage percentages of the interval estimates are close to the nominal values for all parameters in all cases, except for the intervals for λ when the true value of λ satisfies $|\lambda| \geq 1$ in the case in which $n = 100$.

Table 2: Simulation study for the BSCNBS model (continuation).

Case	Parameter	$n = 100$				$n = 250$				$n = 500$			
		AB	SE_1	SE_2	CP	AB	SE_1	SE_2	CP	AB	SE_1	SE_2	CP
$\lambda = 0.25$	α_1	−0.012	0.070	0.073	0.926	−0.003	0.045	0.045	0.944	−0.003	0.032	0.030	0.957
	α_2	−0.010	0.070	0.068	0.949	−0.003	0.045	0.046	0.934	−0.003	0.032	0.031	0.947
	β_1	0.007	0.087	0.087	0.947	0.002	0.056	0.054	0.955	−0.001	0.039	0.040	0.944
	β_2	0.003	0.087	0.091	0.940	0.002	0.056	0.055	0.958	0.000	0.039	0.039	0.955
	λ	0.078	0.247	0.356	0.944	0.013	0.126	0.126	0.952	0.011	0.088	0.091	0.950
$\lambda = 0.50$	α_1	−0.006	0.070	0.072	0.928	−0.005	0.045	0.044	0.954	−0.001	0.032	0.032	0.954
	α_2	−0.007	0.070	0.070	0.933	−0.002	0.045	0.045	0.939	−0.001	0.032	0.032	0.949
	β_1	0.009	0.087	0.085	0.961	0.003	0.055	0.054	0.953	0.001	0.039	0.041	0.940
	β_2	0.006	0.087	0.090	0.942	0.003	0.055	0.056	0.949	0.001	0.039	0.039	0.947
	λ	0.103	0.374	0.469	0.943	0.030	0.199	0.209	0.949	0.016	0.137	0.140	0.947
$\lambda = 0.75$	α_1	−0.008	0.070	0.070	0.940	−0.002	0.045	0.044	0.946	−0.002	0.032	0.031	0.961
	α_2	−0.010	0.070	0.071	0.937	0.000	0.045	0.046	0.945	−0.002	0.032	0.031	0.938
	β_1	0.003	0.085	0.089	0.941	0.004	0.054	0.055	0.940	0.003	0.038	0.039	0.946
	β_2	0.004	0.085	0.088	0.949	0.001	0.054	0.056	0.941	0.001	0.038	0.039	0.947
	λ	0.171	0.568	0.775	0.916	0.041	0.284	0.306	0.939	0.026	0.192	0.196	0.947
$\lambda = 1.00$	α_1	−0.008	0.070	0.071	0.932	−0.005	0.045	0.044	0.946	−0.001	0.032	0.033	0.928
	α_2	−0.004	0.071	0.074	0.935	−0.003	0.045	0.046	0.937	−0.002	0.032	0.032	0.942
	β_1	0.002	0.084	0.088	0.931	0.001	0.053	0.056	0.936	0.002	0.038	0.038	0.952
	β_2	0.005	0.084	0.085	0.950	0.002	0.053	0.054	0.956	0.002	0.038	0.038	0.940
	λ	0.346	0.847	1.526	0.908	0.080	0.379	0.462	0.933	0.032	0.250	0.265	0.945
$\lambda = 2.00$	α_1	−0.009	0.070	0.070	0.949	−0.005	0.045	0.045	0.948	−0.001	0.032	0.032	0.941
	α_2	−0.006	0.070	0.071	0.937	−0.002	0.045	0.045	0.943	−0.002	0.032	0.032	0.940
	β_1	0.004	0.078	0.082	0.933	0.003	0.050	0.054	0.930	0.000	0.035	0.034	0.951
	β_2	0.003	0.079	0.084	0.923	0.000	0.050	0.051	0.939	0.000	0.035	0.037	0.940
	λ	0.875	2.158	3.217	0.886	0.229	0.841	1.038	0.939	0.085	0.531	0.602	0.947
$\lambda = 5.00$	α_1	−0.006	0.071	0.072	0.935	−0.005	0.045	0.046	0.940	−0.001	0.032	0.032	0.944
	α_2	−0.006	0.071	0.073	0.928	−0.002	0.045	0.045	0.952	−0.002	0.032	0.032	0.940
	β_1	0.005	0.067	0.073	0.909	−0.001	0.042	0.044	0.940	0.000	0.029	0.030	0.937
	β_2	0.005	0.068	0.074	0.915	0.000	0.042	0.045	0.920	0.000	0.029	0.032	0.929
	λ	1.668	5.676	5.215	0.873	1.184	3.227	3.772	0.919	0.675	1.839	2.071	0.949

4.3. Real data set: Ais data set

The *ais* data set (see DAAG package, Maindonald and Braun, [29]) includes information about 13 characteristics measured in 202 Australian athletes. We considered two of those variables: the red blood cell count (`rcc`) and the lean body mass (in kg, `lbm`). We model such variables jointly with the BSCNBS distribution. From the data we obtain $s_1 = 4.7186, s_2 = 64.8737, r_1 = 4.6753$ and $r_2 = 62.2709$, providing the following initial values for the estimation algorithm: $\widehat{\alpha}_1^{(0)} = 0.0961, \widehat{\alpha}_2^{(0)} = 0.2034, \widehat{\beta}_1^{(0)} = 4.6969, \widehat{\beta}_2^{(0)} = 63.5590$ and $\lambda^{(0)} = 4.53$. Table 3 shows the estimates for the three considered models. We also use two tests to verify the improved performance of the BSCNBS model compared to the BSNBS and BVBS models.

Table 3: ML estimates for BSCNBS and BSNBS models in betaplasma data set (standard errors in parenthesis).

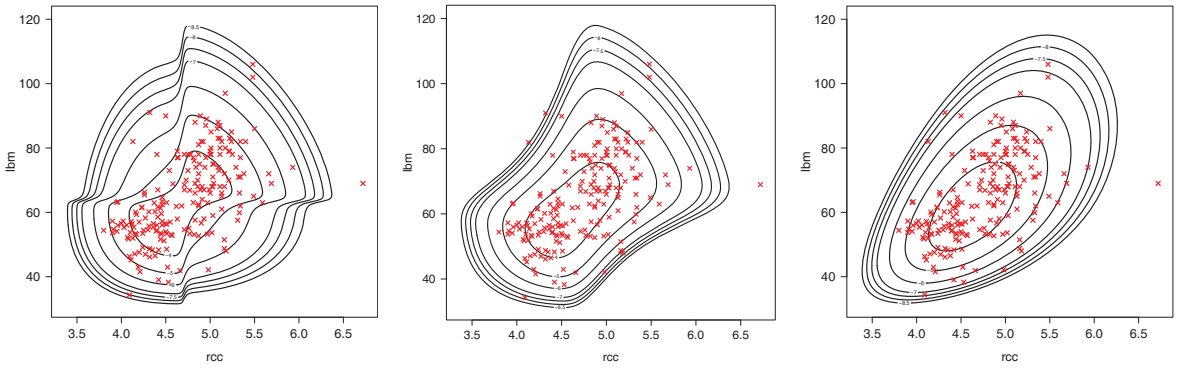
parameter	BSCNBS	BSNBS	BVBS
α_1	0.0962 (0.0048)	0.0961 (0.0048)	0.0962 (0.0048)
α_2	0.2034 (0.0101)	0.2035 (0.0101)	0.2034 (0.0101)
β_1	4.6969 (0.0279)	4.6946 (0.0277)	4.6967 (0.0317)
β_2	63.5588 (0.6720)	63.2352 (0.7882)	63.5568 (0.9049)
λ	4.5003 (1.8784)	1.2132 (0.1843)	—
ρ	—	—	0.5573 (0.0485)
log-likelihood	−881.3825	−885.485	−890.9184
AIC	1772.77	1780.97	1791.84
BIC	1789.31	1797.51	1808.38

The Kolmogorov–Smirnov (KS) statistics are used to verify marginally the BS fit of the `rcc` and `lbm` variates. In addition, it is very important to also consider the bivariate fit of the data to the model. For this, we use an empirical goodness-of-fit test for multivariate distributions proposed in McAssey [31]. We denote A_T as the statistic for such test. Note that both the Akaike information criterion (AIC) (Akaike [1]) and Bayesian information criterion (BIC) (Schwarz [39]) are lower for the BSCNBS model. Additionally, Table 4 shows that both, marginal and bivariate tests provides greater p-values for the BSCNBS model. Thus both, marginal and bivariate tests, suggest better performance for the BSCNBS model.

Table 4: Goodness-of-fit to betaplasma data set (p-values in parenthesis).

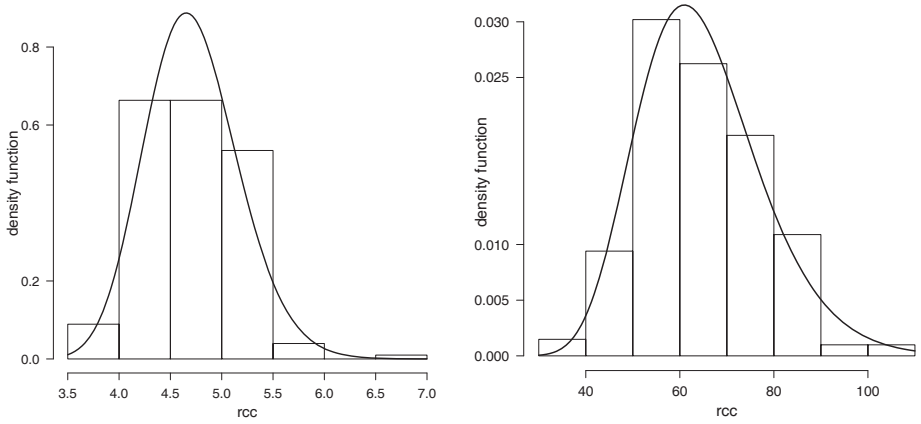
Test	BSCNBS		BSNBS		BVBS	
	rcc	lbm	rcc	lbm	rcc	lbm
KS (marginal)	0.078 (0.172)	0.060 (0.456)	0.079 (0.152)	0.066 (0.334)	0.078 (0.170)	0.060 (0.455)
A_T (bivariate)	5.723 (0.150)		6.574 (0.021)		6.158 (0.063)	

Figure 3 also shows the scatterplot for this data set superimposed on contours of the three fitted models and Figure 4 shows the histogram and estimated density function based on the marginal BS for `rcc` and `lbm` variables. A visual inspection indicates a somewhat better fit of the BSCNSBS relative to the BSNBS model and that the BS distribution is appropriate marginally for this data set. Figure 5 presents the MQR for both variables, the BQR and the respective p-values for the KS test to check the normality hypothesis. Note that, under the usual significance levels, the hypothesis for both, marginal and bivariate residuals, is not rejected, reinforcing the idea that the BSCNBS is appropriate for this data set.



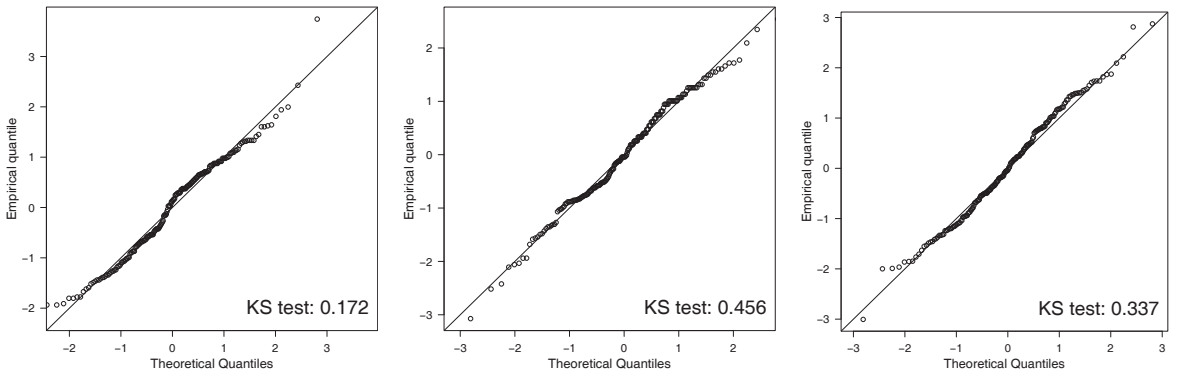
(a) **(b)** **(c)**

Figure 3: Scatterplot of rcc versus lbm for ais data set: (a) BSCNBS; (b) BSNBS and; (c) BVBS models.



(a) **(b)**

Figure 4: Histogram and density function for: (a) `rcc`; and (b) `lbm`, and their estimated density function based on the marginal BS distributions.



(a) **(b)** **(c)**

Figure 5: QQ-plot for the MQR for: (a) the variable `rcc`; (b) the `lbm` variable and; (c) the BQR, based on the fitted BSCNBS model in the ais data set. Also is presented the p-value for the KS test to check if the residuals have standard normal distribution.

5. MULTIVARIATE EXTENSIONS

To obtain a k-dimensional extension of the model discussed in this paper, it is only necessary to identify a specific k-dimensional skewed distribution with normal marginals and skew-normal conditionals to which appropriate marginal transformations are to be applied. For example, one might consider the following joint density.

$$f_1(x_1, x_2, ..., x_k; \lambda) = 2\left[\prod_{i=1}^{k} \phi(x_i)\right] \Phi\left(\frac{\lambda \prod_{i=1}^{k} x_i}{\sqrt{1 + \lambda^2 \prod_{i=1}^{k} x_i^2}}\right), \tag{5.1}$$

which, when marginally transformed, yields a natural extension of the bivariate model discussed in the present paper. Instead of (5.1) we might consider

$$f_2(x_1, x_2, ..., x_k; \lambda) = 2\left[\prod_{i=1}^{k} \phi(x_i)\right] \Phi\left(\lambda \prod_{i=1}^{k} x_i\right), \tag{5.2}$$

which yields the k-dimensional version of the Lemonte *et al.* [28] model.

Both of these models suffer from the fact that only a single dependence parameter, λ, is present. Based on our experience with multivariate normal models and their close relatives, we might prefer to have perhaps $k(k-1)/2$ dependence parameters, if not more, to ensure sufficient flexibility of the model. An extreme example is one which involves use of a k-dimensional joint density which has skew generalized normal conditionals with 2^{k-1} or 3^{k-1} parameters, which is to be transformed to have BS marginals. In practice, some intermediate configuration of dependence parameters might be expected to be appropriate in a particular data setting.

6. CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

The model (2.9) that has been investigated in this paper is, of course, only one of the many bivariate models with BS marginals. In complete generality, one could consider two BS quantile functions and apply them to any copula (i.e., any distribution with standard uniform marginals). Perusal of Nelsen [32] will reveal that essentially there are a limited number of copulas with analytic forms that are readily available for such constructions. Moreover many of the well known copula families have only a single dependence parameter, as is the case with the families of distributions discussed in the present paper. It does thus seem reasonable to consider some of these copula based bivariate BS models as competitors to the models of this paper.

Another approach that might be considered for data sets with BS characteristics, is to take advantage of the fact that the family of univariate BS distributions is an exponential family. Following Arnold and Strauss [6] we might wish to consider the exponential family of bivariate densities with BS conditionals (rather than marginals) as competitors of the models in this paper. Such models have been investigated in Arnold *et al.* [5].

Yet a third general class of models might be considered. For it, assume that $X \sim \text{BS}(\alpha, \beta)$ and that, for each $x > 0$, $Y|X = x \sim \text{BS}(A(x;\underline{\theta}), B(x;\underline{\theta}))$, for certain positive functions $A(x;\underline{\theta})$ and $B(x;\underline{\theta})$. Filus and Filus [21] have investigated models of this genre, in cases in which the roles of the BS distributions are played by normal or exponential distributions.

Multivariate extensions of all the concepts alluded to in this Discussion are readily envisioned.

A. APPENDIX

The Fisher information matrix for the BSCNBS model is given by

$$I(\boldsymbol{\psi}) = \begin{pmatrix} I_{\alpha_1\alpha_1} & I_{\alpha_1\alpha_2} & I_{\alpha_1\beta_1} & I_{\alpha_1\beta_2} & I_{\alpha_1\lambda} \\ I_{\alpha_2\alpha_1} & I_{\alpha_2\alpha_2} & I_{\alpha_2\beta_1} & I_{\alpha_2\beta_2} & I_{\alpha_2\lambda} \\ I_{\beta_1\alpha_1} & I_{\beta_1\alpha_2} & I_{\beta_1\beta_1} & I_{\beta_1\beta_2} & I_{\beta_1\lambda} \\ I_{\beta_2\alpha_1} & I_{\beta_2\alpha_2} & I_{\beta_2\beta_1} & I_{\beta_2\beta_2} & I_{\beta_2\lambda} \\ I_{\lambda\alpha_1} & I_{\lambda\alpha_2} & I_{\lambda\beta_1} & I_{\lambda\beta_2} & I_{\lambda\lambda} \end{pmatrix},$$

with

$$\begin{aligned}
I_{\alpha_1\alpha_1} &= \frac{2}{\alpha_1^2} + \frac{\lambda^3}{\alpha_1^2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{7/2}}\right) + \frac{\lambda^2}{\alpha_1^2}\mathbb{E}\left(\frac{\omega^2 a_1^2 a_2^2}{(1+\lambda^2 a_1^2 a_2^2)^{3}}\right) + \frac{3\lambda^3}{\alpha_1^2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
&\quad - \frac{2\lambda}{\alpha_1^2}\mathbb{E}\left(\frac{\omega a_1 a_2}{(1+\lambda^2 a_1^2 a_2^2)^{3/2}}\right) \\
I_{\alpha_1\alpha_2} &= \frac{\lambda^3}{\alpha_1\alpha_2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{7/2}}\right) + \frac{\lambda^2}{\alpha_1\alpha_2}\mathbb{E}\left(\frac{\omega^2 a_1^2 a_2^2}{(1+\lambda^2 a_1^2 a_2^2)^{3}}\right) - \frac{\lambda}{\alpha_1\alpha_2}\mathbb{E}\left(\frac{\omega a_1 a_2}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
&\quad + \frac{2\lambda^3}{\alpha_1\alpha_2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
I_{\alpha_1\beta_1} &= \frac{2\lambda^3}{\alpha_1^2\beta_1}\mathbb{E}\left(\frac{\omega a_1^2 a_2^3 d_1}{(1+\lambda^2 a_1^2 a_2^2)^{7/2}}\right) + \frac{6\lambda^3}{\alpha_1^2\beta_1}\mathbb{E}\left(\frac{\omega a_1^2 a_2^3 d_1}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) - \frac{\lambda}{2\alpha_1^2\beta_1}\mathbb{E}\left(\frac{\omega a_2 d_1}{(1+\lambda^2 a_1^2 a_2^2)^{3/2}}\right) \\
I_{\alpha_1\beta_2} &= \frac{\lambda^3}{2\alpha_1\alpha_2\beta_2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^2 d_2}{(1+\lambda^2 a_1^2 a_2^2)^{3}}\right) - \frac{\lambda}{2\alpha_1\alpha_2\beta_2}\mathbb{E}\left(\frac{\omega}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
&\quad + \frac{\lambda^3}{\alpha_1\alpha_2\beta_1}\mathbb{E}\left(\frac{\omega a_1^2 a_2^2 d_2}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
I_{\alpha_1\lambda} &= \frac{\lambda^2}{\alpha_1}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{7/2}}\right) - \frac{\lambda}{\alpha_1}\mathbb{E}\left(\frac{\omega^2 a_1^2 a_2^2}{(1+\lambda^2 a_1^2 a_2^2)^{3}}\right) + \frac{1}{\alpha_1}\mathbb{E}\left(\frac{\omega a_1 a_2}{(1+\lambda^2 a_1^2 a_2^2)^{3/2}}\right) \\
I_{\alpha_2\alpha_2} &= \frac{2}{\alpha_2^2} + \frac{\lambda^3}{\alpha_2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{7/2}}\right) + \frac{\lambda^2}{\alpha_2}\mathbb{E}\left(\frac{\omega^2 a_1^2 a_2^2}{(1+\lambda^2 a_1^2 a_2^2)^{3}}\right) + \frac{3\lambda^3}{\alpha_2^2}\mathbb{E}\left(\frac{\omega a_1^3 a_2^3}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
&\quad - \frac{2\lambda}{\alpha_2^2}\mathbb{E}\left(\frac{\omega a_1 a_2}{(1+\lambda^2 a_1^2 a_2^2)^{3/2}}\right) \\
I_{\alpha_2\beta_1} &= \frac{\lambda^3}{2\alpha_1\alpha_2\beta_1}\mathbb{E}\left(\frac{\omega a_1^2 a_2^3 d_1}{(1+\lambda^2 a_1^2 a_2^2)^{7/2}}\right) - \frac{\lambda}{2\alpha_1\alpha_2\beta_1}\mathbb{E}\left(\frac{\omega a_2 d_1}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right) \\
&\quad + \frac{\lambda^3}{\alpha_1\alpha_2\beta_1}\mathbb{E}\left(\frac{\omega a_1^2 a_2^2 d_1}{(1+\lambda^2 a_1^2 a_2^2)^{5/2}}\right)
\end{aligned}$$

$$
\begin{aligned}
I_{\alpha_2\beta_2} &= \frac{\lambda^3}{2\alpha_2^2\beta_2}\mathbb{E}\bigg(\frac{\omega a_1^3a_2^2d_2}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)+\frac{3\lambda^3}{2\alpha_2^2\beta_2}\mathbb{E}\bigg(\frac{\omega a_1^3a_2^2d_2}{(1+\lambda^2a_1^2a_2^2)^{5/2}}\bigg)\\
&\quad-\frac{\lambda}{2\alpha_2^2\beta_2}\mathbb{E}\bigg(\frac{\omega a_1d_2}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
I_{\alpha_2\lambda} &= \frac{\lambda^2}{\alpha_2}\mathbb{E}\bigg(\frac{\omega a_1^3a_2^3}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)-\frac{\lambda}{\alpha_2}\mathbb{E}\bigg(\frac{\omega^2a_1^2a_2^2}{(1+\lambda^2a_1^2a_2^2)^{3}}\bigg)+\frac{1}{\alpha_2}\mathbb{E}\bigg(\frac{\omega a_1a_2}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
I_{\beta_1\beta_1} &= \frac{1}{\alpha_1^2\beta_1^2}+\mathbb{E}[(T_1+\beta_1)^{-2}]+\frac{\lambda^3}{4\alpha_1^2\beta_1^2}\mathbb{E}\bigg(\frac{\omega a_1a_2^3d_1^2}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)+\frac{\lambda^2}{4\alpha_1^2\beta_1^2}\mathbb{E}\bigg(\frac{\omega^2a_2^2d_1^2}{(1+\lambda^2a_1^2a_2^2)^{3}}\bigg)\\
&\quad+\frac{3\lambda^3}{4\alpha_1^2\beta_1^2}\mathbb{E}\bigg(\frac{\omega a_1a_2^3d_1^2}{(1+\lambda^2a_1^2a_2^2)^{5/2}}\bigg)-\frac{\lambda}{4\alpha_1^2\beta_1^2}\mathbb{E}\bigg(\frac{\omega a_2d_1}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
&\quad-\frac{\lambda}{2\alpha_1^2\beta^{5/2}}\mathbb{E}\bigg(\frac{\omega a_2T_1^{1/2}}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
I_{\beta_1\beta_2} &= \frac{\lambda^3}{4\alpha_1\alpha_2\beta_1\beta_2}\mathbb{E}\bigg(\frac{\omega a_1^2a_2^2d_1d_2}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)-\frac{\lambda}{4\alpha_1\alpha_2\beta_1\beta_2}\mathbb{E}\bigg(\frac{\omega d_1d_2}{(1+\lambda^2a_1^2a_2^2)^{5/2}}\bigg)\\
&\quad+\frac{\lambda^3}{2\alpha_1\alpha_2\beta_1\beta_2}\mathbb{E}\bigg(\frac{\omega a_1^2a_2^2d_1d_2}{(1+\lambda^2a_1^2a_2^2)^{5/2}}\bigg)\\
I_{\beta_1\lambda} &= \frac{\lambda^2}{2\alpha_1\beta_1}\mathbb{E}\bigg(\frac{\omega a_1^2a_2^3d_1}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)+\frac{1}{2\alpha_1\beta_1}\mathbb{E}\bigg(\frac{\omega a_2d_1}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
I_{\beta_2\beta_2} &= \frac{1}{\alpha_2^2\beta_2^2}+\mathbb{E}[(T_2+\beta_2)^{-2}]+\frac{\lambda^3}{4\alpha_2\beta_2^2}\mathbb{E}\bigg(\frac{\omega a_1^3a_2d_2^2}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)+\frac{\lambda^2}{4\alpha_2\beta_2^2}\mathbb{E}\bigg(\frac{\omega^2a_1^2d_2^2}{(1+\lambda^2a_1^2a_2^2)^{3}}\bigg)\\
&\quad+\frac{3\lambda^3}{4\alpha_2\beta_2^2}\mathbb{E}\bigg(\frac{\omega a_1^3a_2d_2^2}{(1+\lambda^2a_1^2a_2^2)^{5/2}}\bigg)-\frac{\lambda}{4\alpha_2^2\beta_2^2}\mathbb{E}\bigg(\frac{\omega a_1d_2}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
&\quad-\frac{\lambda}{2\alpha_2^2\beta_2^{5/2}}\mathbb{E}\bigg(\frac{\omega a_1T_2^{1/2}}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
I_{\beta_2\lambda} &= \frac{\lambda^2}{2\alpha_2\beta_2}\mathbb{E}\bigg(\frac{\omega a_1^3a_2^2d_2}{(1+\lambda^2a_1^2a_2^2)^{7/2}}\bigg)+\frac{1}{2\alpha_2\beta_2}\mathbb{E}\bigg(\frac{\omega a_1d_2}{(1+\lambda^2a_1^2a_2^2)^{3/2}}\bigg)\\
I_{\lambda\lambda} &= \mathbb{E}\bigg(\frac{\omega^2a_1^2a_2^2}{(1+\lambda^2a_1^2a_2^2)^{3}}\bigg)+3\lambda\mathbb{E}\bigg(\frac{\omega a_1^3a_2^3}{(1+\lambda^2a_1^2a_2^2)^{3}}\bigg),
\end{aligned}
$$

where $d_j=(T_j/\beta_j)^{1/2}+(\beta_j/T_j)^{1/2}$, $j=1,2$, $\omega=\phi(b)/\Phi(b)$ and $b=\lambda a_1a_2/\sqrt{1+\lambda a_1^2a_2^2}$. Note that for $\boldsymbol{\psi}_0=(\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda=0)$, this matrix is reduced to

$$
I(\boldsymbol{\psi}_0)=\operatorname{diag}\bigg(\frac{2}{\alpha_1^2},\frac{2}{\alpha_2^2},\frac{1}{\alpha_1^2\beta_1^2}+\mathbb{E}[(T_1+\beta_1)^{-2}],\frac{1}{\alpha_2^2\beta_2^2}+\mathbb{E}[(T_2+\beta_2)^{-2}],\frac{1}{2\sqrt{2\pi}}\bigg).
$$

Then, the determinant of $I(\boldsymbol{\psi}_0)$ is

$$
|I(\boldsymbol{\psi}_0)|=\frac{2}{\sqrt{2\pi}\alpha_1^2\alpha_2^2}\prod_{j=1}^{2}\big\{(\alpha_j\beta_j)^{-2}+\mathbb{E}\big[(T_j+\beta_j)^{-2}\big]\big\}>0.
$$

Therefore, the Fisher information matrix is not singular at $\lambda=0$.

ABBREVIATIONS

The following abbreviations are used in this manuscript:

BS	Birnbaum–Saunders
SN	Skew-normal
SGN	Skew-generalized-normal
SCN	Skew-curved-normal
SCNBS	Skew-curved-normal-Birnbaum–Saunders
BSCNBS	Bivariate skew-curved-normal-Birnbaum–Saunders
BVBS	Bivariate Birnbaum–Saunders
SNBS	Skew-normal-Birnbaum–Saunders
BSNBS	Bivariate skew-normal-Birnbaum–Saunders
FIM	Fisher information matrix
ML	Maximum likelihood
MQR	Marginal quantile residuals
BQR	Bivariate quantile residuals
AB	Absolute bias
SE_1	Mean of the standard errors
SE_2	Standard deviation of the estimated parameters
CP	Coverage proportion
KS	Kolmogorov–Smirnov
AIC	Akaike information criterion
BIC	Bayesian information criterion

ACKNOWLEDGMENTS

The authors would like to thank the associate editor and the anonymous referees for their comments and suggestions, which significantly improved our manuscript. The research of H. W. Gómez was supported by SEMILLERO UA-2023 (Chile).

REFERENCES

[1] Akaike, H. (1974). A new look at statistical model identification, *IEEE Transaction on Automatic Control*, **19**, 716–723.

[2] Arellano-Valle, R.B.; Gómez, H.W. and Quintana, F.A. (2004). A new class of skew-normal distributions, *Communication in Statistics – Theory and Methods*, **33**, 1465–1480.

[3] Arnold, B.C.; Castillo, E. and Sarabia, J.M. (2002). Conditionally specified multivariate skewed distributions, *Sankhya Series A*, **64**, 206–226.

[4] Arnold, B.C.; Castillo, E. and Sarabia, J.M. (2007). Distributions with generalized skewed conditionals and mixtures of such distributions, *Communication in Statistics – Theory and Methods*, **36**, 1493–1503.

[5] Arnold, B.C.; Gallardo, D.I. and Gómez, H.W. (2021). A note on the Birnbaum–Saunders conditionals model, *Symmetry*, **13**, 762.

[6] Arnold, B.C. and Strauss, D.J. (1991). Bivariate distributions with conditionals in prescribed exponential families, *Journal of the Royal Statistical Society: Series B*, **53**, 365–375.

[7] Arrué, J.; Arellano-Valle, R.B.; Gómez, H.W. and Leiva, V. (2020). On a new type of Birnbaum–Saunders models and its inference and application to fatigue data, *Journal of Applied Statistics*, **47**(13-14), 2690–2710.

[8] Athayde, E. (2017). A characterization of the generalized Birnbaum–Saunders distribution, *REVSTAT*, **15**(3), 333–354.

[9] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[10] Balakrishnan, N. and Alam, F.M.A. (2022). Maximum likelihood estimation of the parameters of student's t Birnbaum–Saunders distribution: a comparative study, *Communications in Statistics – Simulation and Computation*, **51**(3), 793–822.

[11] Balakrishnan, N.; Gupta, R.C.; Kundu, D.; Leiva, V. and Sanhueza, A. (2011). On some mixture models based on the Birnbaum–Saunders distribution and associated inference, *Journal of Statistical Planning and Inference*, **141**, 2175–2190.

[12] Balakrishnan, N.; Leiva, V.; Sanhueza, A. and Vilca, F. (2009). Estimation in the Birnbaum–Saunders distribution based on scale-mixture of normals and the EM-algorithm, *SORT*, **33**, 171–192.

[13] Balakrishnan, N. and Kundu, D. (2019). Birnbaum–Saunders distribution: a review of models, analysis, and applications, *Applied Stochastic Models in Business and Industry*, **35**, 4–49.

[14] Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions, *J. Appl. Probab.*, **6**, 319–327.

[15] Bourguignon, M.; Leao, J.; Leiva, V. and Santos-Neto, M. (2017). The transmuted Birnbaum–Saunders distribution, *REVSTAT*, **15**(4), 601–628.

[16] Carrasco, J.M.F.; Figueroa-Zuñiga, J.I.; Leiva, V.; Riquelme, M. and Aykroyd, R.G. (2020). An errors-in-variables model based on the Birnbaum–Saunders distribution and its diagnostics with an application to earthquake data, *Stochastic Environmental Research and Risk Assessment*, **34**, 369–380.

[17] Dasilva, A.; Dias, R.; Leiva, V.; Marchant, C. and Saulo, H. (2020). [Invited tutorial] Birnbaum–Saunders regression models: a comparative evaluation of three approaches, *Journal of Statistical Computation and Simulation*, **90**(14), 2552–2570.

[18] Dempster, A.P.; Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society B*, **39**, 1–38.

[19] Dunn, P.K. and Smyth, G.K. (1996). Randomized quantile residuals, *Journal of Computational and Graphical Statistics*, **5**, 236–244.

[20] Faraway, J.; Marsaglia, G.; Marsaglia, J. and Baddeley, A. (2019). *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions*, R package version 1.2-2.

[21] Filus, J.K. and Filus, L.Z. (2013). A method for multivariate probability distributions construction via parameter dependence, *Communication in Statistics – Theory and Methods*, **42**, 716–721.

[22] Genz, A.; Bretz, F.; Miwa, T.; Mi, X.; Leisch, F.; Scheipl, F. and Hothorn, T. (2020). *mvtnorm: Multivariate Normal and t Distributions*, R package version 1.1-1.

[23] Gómez, H.W.; Castro, L.M.; Salinas, H.S. and Bolfarine, H. (2010). Properties and inference on the skew-curved-symmetric family of distributions, *Communication in Statistics – Theory and Methods*, **39**, 884–898.

[24] Kalliovirta, L. (2008). Quantile residuals for multivariate models, *Discussion papers / Helsinki Center of Economic Research; No. 247*, Helsinki: Helsinki Center of Economic Research.

[25] Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione, *Giornale dell Instituto Italiano degli Attuari*, **4**, 83–91.

[26] Kundu, D.; Balakrishnan, N. and Jamalizadeh, A. (2010). Bivariate Birnbaum–Saunders distribution and associated inference, *Journal of Multivariate Analysis*, **101**, 113–125.

[27] Leiva, V. (2016). *The Birnbaum–Saunders Distribution*, Academic Press, New York.

[28] Lemonte, A.J.; Martínez-Flóres, G. and Moreno-Arenas, G. (2015). Multivariate Birnbaum–Saunders distribution: properties and associated inference, *Journal of Statistical Computation and Simulation*, **85**, 374–392.

[29] Maindonald, J.H. and Braun, W.J. (2020). *DAAG: Data Analysis and Graphics Data and Functions*, R package version 1.24.

[30] Martínez-Flórez, G.; Bolfarine, H.; Gómez, Y.M. and Gómez, H.W. (2020). A unification of families of Birnbaum–Saunders distributions with applications, *REVSTAT*, **18**(5), 637–660.

[31] McAssey, M.P. (2013). An empirical goodness-of-fit test for multivariate distributions, *Journal of Applied Statistics*, **40**, 1120–1131.

[32] Nelsen, R.B. (2006). *An Introduction to Copulas*, Second Edition, Springer, New York.

[33] Ng, H.K.T.; Kundu, D. and Balakrishnan, N. (2003). Modified moment estimation for the two parameter Birnbaum–Saunders distribution, *Comput. Statist. Data Anal.*, **43**, 283–298.

[34] Pradhan, B. and Kundu, D. (2013). Inference and optimal censoring schemes for progressively censored Birnbaum–Saunders distribution *Journal of Statistical Planning and Inference*, **143**, 1098–1108.

[35] R Development Core Team R. (2020). *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

[36] Reyes, J.; Barranco-Chamorro, I.; Gallardo, D.I. and Gómez, H.W. (2018). Generalized modified slash Birnbaum–Saunders distribution, *Symmetry*, **10**, 724.

[37] Reyes, J.; Vilca, F.; Gallardo, D.I. and Gómez, H.W. (2017). Modified slash Birnbaum–Saunders distribution, *Hacettepe Journal of Mathematics and Statistics*, **46**, 969–984.

[38] Romeiro, R.G.; Vilca, F. and Balakrishnan, N. (2018). A robust multivariate Birnbaum–Saunders distribution: EM estimation, *Statistics*, **52**, 321–344.

[39] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.

[40] Yee, T.W. (2020). *VGAM: Vector Generalized Linear and Additive Models*, R package version 1.1-3.

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 21–38
https://doi.org/10.57805/revstat.v21i1.395

Folded Bivariate Distributions as Models for Magnitude Correlation

Authors: Emmanuel Afuecheta
– Department of Mathematics and Statistics,
King Fahd University of Petroleum & Minerals,
Dhahran, Saudi Arabia
emmanuel.afuecheta@kfupm.edu.sa

Saralees Nadarajah
– Department of Mathematics, University of Manchester,
Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

Stephen Chan ✉
– Department of Mathematics and Statistics, American University of Sharjah,
UAE
schan@aus.edu

Received: January 2020 Revised: June 2021 Accepted: June 2021

Abstract:

- The concept of *magnitude correlation* requires the use of folded bivariate distributions. However, apart from the folded bivariate normal and folded bivariate t distributions (of these two only the former has received any real applications), nothing is known about folded bivariate distributions. Here, we introduce six new folded bivariate distributions. Applications involving stock indices of ten major economies show the value of the proposed distributions.

Keywords:

- *estimation; magnitude correlation of stock returns; value at risk.*

AMS Subject Classification:

- 62E15.

✉ Corresponding author.

1. INTRODUCTION

Let X and Y be two random variables taking values on the entire real line (for example, stock returns for two different commodities). Their *magnitude correlation* is defined as the correlation between $|X|$ and $|Y|$, i.e., the correlation between the absolute values of X and Y. Its derivation clearly requires the use of folded bivariate distributions.

The concept of magnitude correlation arises in many areas of the sciences, engineering and medicine. One prominent area is stock modeling. Let X denote the stock return for one commodity and Y the stock return for another commodity. It is of interest to know if large values of X in magnitude are associated with large values of Y in magnitude, if large values of X in magnitude are associated with small values of Y in magnitude, if small values of X in magnitude are associated with large values of Y in magnitude, or if small values of X in magnitude are associated with small values of Y in magnitude. The magnitude correlation of stock returns has been studied by many authors. Some recent examples are: Firth [22] observes that the "correlation between the magnitudes of the changes in dividends and changes in future earnings (over the next two years) is highly significant and positive"; While modeling volatility of Dhaka stock exchange, Islam *et al.* [32] observe that the "correlations between the magnitudes of returns on nearby days are positive and statistically significant"; While investigating the overnight effect on the Taiwan stock market, Tsai *et al.* [49] examine the "cross correlation between the magnitude of daytime (trading hour) returns, overnight (off-hour session) returns, and total (close-to-close) returns"; Tsai *et al.* [49] observe "a larger magnitude of overnight return implies a higher probability that the sign of the following daytime return is the opposite of the sign of overnight return"; Chabakauri [11] finds "a positive relationship between the amount of leverage in the economy and magnitudes of stock return correlations and volatilities"; Fukuda and Tanaka [23] show that "during the global financial crisis the magnitude correlation between TIBOR and LIBOR reversed depending on whether they were in yen or dollars"; Bhamra *et al.* [6] show that the magnitude of asset return correlation depends on "three structural parameters: the degree of market integration, the level of fundamental correlation and the rate of time preference"; While investigating equity market reactions to CreditWatch events, Gu *et al.* [25] find evidence of a "positive correlation between the magnitude of the cumulative abnormal returns prior to the listing day and the magnitude of the rating changes announced on the delisting day"; Hamalainen [28] says "Correlation between the magnitudes of asset returns is an overlooked concept in financial research. It affects portfolio variance explicitly when the directions of returns are predictable". See also Kutergin and Filimonov [39].

Magnitude correlation is one area requiring the use of folded bivariate and folded multivariate distributions. Other areas include: approximations to the mean and variance of the index of dissimilarity in contingency tables (Inman and Bradley [31], Mulekar *et al.* [41]); noise sensitivity of a new singularity index (Muralidhar *et al.* [42]); distribution and estimation of trading costs (Kourtis [38]); the joint distribution of indemnity payment and allocated loss adjustment expense for general liability claims (Guillou *et al.* [26]).

In all of these applications and others, only the folded bivariate/multivariate normal distribution appears to have been used. We are not aware of applications of any other folded bivariate/multivariate distribution. The normal distribution does not give good fits to many

types of data including heavy tailed data (for example, stock returns). Hence, there is a need for folded bivariate/multivariate distributions for non-normal data.

Psarakis and Panaretos [45] were the first to introduce the folded bivariate normal distribution and study its statistical properties. They derived its marginal distributions and joint moment generating function. Chakraborty and Chatterjee [12] gave a multivariate form of the folded normal distribution and derived expressions for its mean vector, covariance matrix and joint moment generating function. They also discussed possible areas of applications of the folded multivariate normal distribution.

We are aware of only one folded bivariate distribution for non-normal data, the folded bivariate t distribution. We are aware of no folded multivariate distributions for non-normal data. The folded bivariate t distribution was also introduced by Psarakis and Panaretos [45]. They derived its marginal distributions and established its relationship to the folded bivariate normal distribution.

Neither of the papers (Psarakis and Panaretos [45] or Chakraborty and Chatterjee [12]) discussed real data applications or even simulation studies. The aim of this paper is to:

i) introduce six new folded bivariate distributions;
ii) illustrate real data applications of all of the folded bivariate distributions.

The six new folded bivariate distributions are based on the: bivariate skew normal distribution due to Azzalini and Dalla Valle [4]; bivariate skew t distribution due to Azzalini and Capitanio [3]; bivariate logistic distribution due to Gumbel [27]; bivariate Kotz type distribution due to Kotz [36]; bivariate Laplace distribution of the first kind due to Eltoft *et al.* [19]; bivariate Laplace distribution of the second kind due to Ernst [20]. We have chosen these distributions because they are some of the most tractable and applied bivariate distributions for non-normal data, see Balakrishnan and Lai [5] and references therein.

For each of the new distributions and for the two known ones, we give expressions for the joint probability density function. Expressions for the joint cumulative distribution function, joint moment generating function and the log-likelihood function can be obtained from the corresponding author. As a by product of the six new distributions, we also introduce two new univariate distributions: the folded univariate skew normal distribution and the folded univariate skew t distribution.

Our real data application involves forty five bivariate data sets on log returns of stocks. We show that:

i) the folded bivariate t and folded bivariate skew t distributions provide the best fits for the majority of the data sets;
ii) each of the folded distributions outperforms the corresponding truncated unfolded version for each of the forty five data sets.

The latter observation is a further advocate for the need for folded bivariate distributions.

The expressions in Section 2 involve standard normal cumulative distribution function defined by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

and the modified Bessel function of the second kind (Abramowitz and Stegun [1]) defined by

$$K_\nu(x) = \begin{cases} \dfrac{\pi \csc(\pi\nu)}{2}[I_{-\nu}(x) - I_\nu(x)], & \text{if } \nu \notin \mathbb{Z}, \\ \lim\limits_{\mu\to\nu} K_\mu(x), & \text{if } \nu \in \mathbb{Z}, \end{cases}$$

where $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order ν defined by

$$I_\nu(x) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+\nu+1)k!}\left(\frac{x}{2}\right)^{2k+\nu}.$$

2. NEW FOLDED BIVARIATE DISTRIBUTIONS

Let (X, Y) denote a random vector on $(-\infty, \infty) \times (-\infty, \infty)$ with joint probability density function $f_{X,Y}(x, y)$ and joint cumulative distribution function $F_{X,Y}(x, y)$. Set $(U, V) = (|X|, |Y|)$. Then the joint probability density function of (U, V) is

$$f_{U,V}(u, v) = f_{X,Y}(u, v) + f_{X,Y}(u, -v) + f_{X,Y}(-u, v) + f_{X,Y}(-u, -v). \tag{2.1}$$

The distribution given by $f_{U,V}(u, v)$ is said to be the folded version of the distribution given by $f_{X,Y}(x, y)$.

Let f_X, f_Y denote the marginal probability density functions of (X, Y). Let f_U, f_V denote the marginal probability density functions of (U, V). It follows from (2.1) that $f_U(u) = f_X(u) + f_X(-u)$ and $f_V(v) = f_Y(v) + f_Y(-v)$.

2.1. Folded bivariate normal distribution

The bivariate normal distribution due to the work of Laplace, Plana, Gauss and Bravais is given by the joint probability density function

$$f_{X,Y}(x, y) = \frac{1}{2\pi s_1 s_2\sqrt{1-\rho^2}} \exp\left[-\frac{Q(x, y)}{2(1-\rho^2)}\right]$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, where

$$Q(x, y) = \left(\frac{x-\mu_1}{s_1}\right)^2 + \left(\frac{y-\mu_2}{s_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{s_1}\right)\left(\frac{y-\mu_2}{s_2}\right) \tag{2.2}$$

for $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$ and $-1 < \rho < 1$. The corresponding folded version has the joint probability density function

$$f_{U,V}(u, v) = \frac{1}{2\pi s_1 s_2\sqrt{1-\rho^2}}\left\{\exp\left[-\frac{Q(u, v)}{2(1-\rho^2)}\right] + \exp\left[-\frac{Q(u, -v)}{2(1-\rho^2)}\right] + \exp\left[-\frac{Q(-u, v)}{2(1-\rho^2)}\right] + \exp\left[-\frac{Q(-u, -v)}{2(1-\rho^2)}\right]\right\}.$$

The marginals of the folded bivariate normal distribution are the folded univariate normal distributions due to Leone *et al.* [40].

2.2. Folded bivariate t distribution

The bivariate t distribution is given by the joint probability density function

$$f_{X,Y}(x,y) = \frac{\Gamma(1+\nu/2)}{\nu\pi\Gamma(\nu/2)s_1 s_2\sqrt{1-\rho^2}}\left[1+\frac{Q(x,y)}{\nu(1-\rho^2)}\right]^{-1-\nu/2}$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $\nu > 0$ and $-1 < \rho < 1$, where Q is given by (2.2). The corresponding folded version has the joint probability density function

$$\begin{aligned} f_{U,V}(u,v) \;=\; & \frac{\Gamma(1+\nu/2)}{\nu\pi\Gamma(\nu/2)s_1 s_2\sqrt{1-\rho^2}}\left\{\left[1+\frac{Q(u,v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2}\right. \\ & +\left[1+\frac{Q(u,-v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} + \left[1+\frac{Q(-u,v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} \\ & \left.+\left[1+\frac{Q(-u,-v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2}\right\} \end{aligned} \tag{2.3}$$

for $u > 0$, $v > 0$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $\nu > 0$ and $-1 < \rho < 1$. The particular case for $\nu = 1$ is the folded bivariate Cauchy distribution, another new folded bivariate distribution. The marginals of the folded bivariate t distribution are the folded univariate t distributions due to Psarakis and Panaretos [44].

2.3. Folded bivariate skew normal distribution

The bivariate skew normal distribution due to Azzalini and Dalla Valle [4] has the joint probability density function specified by

$$f_{X,Y}(x,y) = \frac{1}{\pi s_1 s_2\sqrt{1-\rho^2}}\exp\left[-\frac{Q(x,y)}{2(1-\rho^2)}\right]\Phi(\alpha_1 x+\alpha_2 y)$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $-\infty < \alpha_1 < \infty$, $-\infty < \alpha_2 < \infty$ and $-1 < \rho < 1$, where Q is given by (2.2). The corresponding folded version has the joint probability density function

$$\begin{aligned} f_{U,V}(u,v) \;=\; & \frac{1}{\pi s_1 s_2\sqrt{1-\rho^2}}\left\{\exp\left[-\frac{Q(u,v)}{2(1-\rho^2)}\right]\Phi(\alpha_1 u+\alpha_2 v)\right. \\ & +\exp\left[-\frac{Q(u,-v)}{2(1-\rho^2)}\right]\Phi(\alpha_1 u-\alpha_2 v) \\ & +\exp\left[-\frac{Q(-u,v)}{2(1-\rho^2)}\right]\Phi(-\alpha_1 u+\alpha_2 v) \\ & \left.+\exp\left[-\frac{Q(-u,-v)}{2(1-\rho^2)}\right]\Phi(-\alpha_1 u-\alpha_2 v)\right\} \end{aligned}$$

for $u > 0$, $v > 0$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $-\infty < \alpha_1 < \infty$, $-\infty < \alpha_2 < \infty$ and $-1 < \rho < 1$. The marginals of the folded bivariate skew normal distribution are the folded univariate skew normal distributions, which appear to be new.

2.4. Folded bivariate skew t distribution

The bivariate skew t distribution due to Azzalini and Capitanio [3] has the joint probability density function specified by

$$f_{X,Y}(x,y) = \frac{2\Gamma(1+\nu/2)}{\nu\pi\Gamma(\nu/2)s_1 s_2\sqrt{1-\rho^2}}\left[1+\frac{Q(x,y)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} R(x,y)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, where Q is given by (2.2) and

$$R(x,y) = G\left(\left(\frac{\alpha_1(x-\mu_1)}{\sqrt{s_{11}}} + \frac{\alpha_2(y-\mu_2)}{\sqrt{s_{22}}}\right)\sqrt{\frac{\nu+2}{Q(x,y)+2}}; \nu+2\right),$$

where $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $-\infty < \alpha_1 < \infty$, $-\infty < \alpha_2 < \infty$, $\nu > 0$, $-1 < \rho < 1$, and $G(\cdot;\nu)$ denotes the cumulative distribution function of a Student's t random variable with ν degrees of freedom. The corresponding folded version has the joint probability density function

$$\begin{aligned} f_{U,V}(u,v) \;=\; & \frac{2\Gamma(1+\nu/2)}{\nu\pi\Gamma(\nu/2)s_1 s_2\sqrt{1-\rho^2}}\left\{\left[1+\frac{Q(u,v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} R(u,v)\right. \\ & +\left[1+\frac{Q(u,-v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} R(u,-v) \\ & +\left[1+\frac{Q(-u,v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} R(-u,v) \\ & \left.+\left[1+\frac{Q(-u,-v)}{\nu(1-\rho^2)}\right]^{-1-\nu/2} R(-u,-v)\right\} \end{aligned}$$

for $u > 0$, $v > 0$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $-\infty < \alpha_1 < \infty$, $-\infty < \alpha_2 < \infty$, $\nu > 0$ and $-1 < \rho < 1$. The marginals of the folded bivariate skew t distribution are the folded univariate skew t distributions which appear to be new too. The G term does admit closed form expressions if ν is an integer, see Jamalizadeh *et al.* [33].

2.5. Folded bivariate logistic distribution

The bivariate logistic distribution due to Gumbel [27] has the joint probability density function specified by

$$f_{X,Y}(x,y) = \frac{2\exp[-(x-\mu_1)/s_1 - (y-\mu_2)/s_2]}{s_1 s_2\{1+\exp[-(x-\mu_1)/s_1] + \exp[-(y-\mu_2)/s_2]\}^3}$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, and $s_2 > 0$. The corresponding folded version has the joint probability density function

$$\begin{aligned} f_{U,V}(u,v) \;=\; & \frac{2}{s_1 s_2}\left\{ \frac{\exp[-(u-\mu_1)/s_1 - (v-\mu_2)/s_2]}{\{1+\exp[-(u-\mu_1)/s_1]+\exp[-(v-\mu_2)/s_2]\}^3} \right. \\ & + \frac{\exp[-(u-\mu_1)/s_1 + (v+\mu_2)/s_2]}{\{1+\exp[-(u-\mu_1)/s_1]+\exp[(v+\mu_2)/s_2]\}^3} \\ & + \frac{\exp[(u+\mu_1)/s_1 - (v-\mu_2)/s_2]}{\{1+\exp[(u+\mu_1)/s_1]+\exp[-(v-\mu_2)/s_2]\}^3} \\ & \left. + \frac{\exp[(u+\mu_1)/s_1 + (v+\mu_2)/s_2]}{\{1+\exp[(u+\mu_1)/s_1]+\exp[(v+\mu_2)/s_2]\}^3} \right\} \end{aligned}$$

for $u > 0$, $v > 0$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, and $s_2 > 0$. The marginals of the folded bivariate logistic distribution are the folded univariate logistic distributions due to Cooray *et al.* [13].

2.6. Folded bivariate Kotz type distribution

The bivariate Kotz type distribution is given by the joint probability density function

$$f_{X,Y}(x,y) = \frac{s r^{N/s}\left(1-\rho^2\right)^{1/2-N}}{s_1 s_2 \pi \Gamma(N/s)} [Q(x,y)]^{N-1} \exp\left\{ -\frac{r}{(1-\rho^2)^s} [Q(x,y)]^s \right\} \tag{2.4}$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $N > 0$, $r > 0$, $s > 0$ and $-1 < \rho < 1$, where Q is given by (2.2). When $s = 1$, this is the original Kotz distribution introduced in Kotz [36]. When $N = 1$, $s = 1$ and $r = 1/2$, (2.4) reduces to the bivariate normal probability density function. When $N = 1$, $s = 1/2$ and $r = 1$, (2.4) reduces to the joint probability density function of the bivariate Laplace distribution of the second kind. The folded bivariate Kotz type distribution corresponding to (2.4) has the joint probability density function

$$\begin{aligned} f_{U,V}(u,v) \;=\; & \frac{s r^{N/s}\left(1-\rho^2\right)^{1/2-N}}{s_1 s_2 \pi \Gamma(N/s)} \left[[Q(u,v)]^{N-1} \exp\left\{ -\frac{r}{(1-\rho^2)^s} [Q(u,v)]^s \right\} \right. \\ & + [Q(u,-v)]^{N-1} \exp\left\{ -\frac{r}{(1-\rho^2)^s} [Q(u,-v)]^s \right\} \\ & + [Q(-u,v)]^{N-1} \exp\left\{ -\frac{r}{(1-\rho^2)^s} [Q(-u,v)]^s \right\} \\ & \left. + [Q(-u,-v)]^{N-1} \exp\left\{ -\frac{r}{(1-\rho^2)^s} [Q(-u,-v)]^s \right\} \right] \end{aligned}$$

for $u > 0$, $v > 0$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $N > 0$, $r > 0$, $s > 0$ and $-1 < \rho < 1$. The marginals of the folded bivariate Kotz type distribution are the folded univariate exponential power distributions due to Nadarajah and Bakar [43].

2.7. Folded bivariate Laplace distribution of the first kind

The bivariate Laplace distribution of the first kind due to Eltoft *et al.* [19] has the joint probability density function specified by

$$f_{X,Y}(x,y) = \frac{1}{\pi r} K_0\left(\sqrt{\frac{2}{r}}\sqrt{Q(x,y)}\right)$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $r > 0$ and $-1 < \rho < 1$, where Q is given by (2.2). The corresponding folded version has the joint probability density function

$$\begin{aligned} f_{U,V}(u,v) \;=\; & \frac{1}{\pi r}\left[K_0\left(\sqrt{\frac{2}{r}}\sqrt{Q(u,v)}\right) + K_0\left(\sqrt{\frac{2}{r}}\sqrt{Q(u,-v)}\right)\right. \\ & \left. + K_0\left(\sqrt{\frac{2}{r}}\sqrt{Q(-u,v)}\right) + K_0\left(\sqrt{\frac{2}{r}}\sqrt{Q(-u,-v)}\right)\right] \end{aligned}$$

for $u > 0$, $v > 0$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $s_1 > 0$, $s_2 > 0$, $r > 0$ and $-1 < \rho < 1$. The marginals of the folded bivariate Laplace distribution of the first kind are the folded univariate Laplace distributions, special cases of the folded exponential power distribution due to Nadarajah and Bakar [43].

2.8. Folded bivariate Laplace distribution of the second kind

The folded bivariate Laplace distribution of the second kind is the particular case of the folded bivariate Kotz type distribution for $N = 1$, $s = 1/2$ and $r = 1$. So, the corresponding joint probability density function follows from the expression given in Section 2.6. The marginals of the folded bivariate Laplace distribution of the second kind are also the folded univariate Laplace distributions.

3. APPLICATION

In this section, we study the magnitude correlation for daily log returns of stock values from the 3rd of January 2000 to the 28th of February 2014 for the ten countries: the United States of America (S & P 500), Canada (S & P TSX), the United Kingdom (FTSE 100), Germany (DAX), China (SSE), Japan (Nikki), Brazil (BOVESPA), Argentina (MERVAL), South Africa (FTSE/JSE) and Nigeria (S & P). The data were obtained from the database Datastream.

The distributions in Section 2 assume that the data on each country are independent and identically distributed (i.e., randomness), have no serial correlation, and have no heteroskedasticity. We tested for randomness using Cox and Stuart [15]'s test, the rank test and the turning point test. We tested for no serial correlation using Durbin and Watson

[16, 17, 18]'s method and the method due to Godfrey [24] and Breusch [8]. We tested for no heteroskedasticity using Breusch and Pagan [9]'s test. The corresponding p-values not reported here showed no evidence are randomness, no serial correlation or no heteroskedasticity.

The magnitude correlation between returns for any two countries can be studied by:

i) fitting a folded bivariate distribution to positive log returns from the countries (that is, considering only those days where the log returns are positive for both countries);

ii) fitting a truncated unfolded bivariate distribution (truncation made to the positive quadrant) to positive log returns from the countries.

We have data for ten countries, so forty five pairs of bivariate data.

Section 2 describes eight folded bivariate distributions. Each of these has a corresponding truncated unfolded version. These sixteen distributions were fitted to positive log returns from: USA / CAD, USA / UK, USA / GER, USA / CHI, USA / JPN, USA / BRA, USA / ARG, USA / SA, USA / NG, CAD / UK, CAD / GER, CAD / CHI, CAD / JPN, CAD / BRA, CAD / ARG, CAD / SA, CAD / NG, UK / GER, UK / CHI, UK / JPN, UK / BRA, UK / ARG, UK / SA, UK / NG, GER / CHI, GER / JPN, GER / BRA, GER / ARG, GER / SA, GER / NG, CHI / JPN, CHI / BRA, CHI / ARG, CHI / SA, CHI / NG, JPN / BRA, JPN / ARG, JPN / SA, JPN / NG, BRA / ARG, BRA / SA, BRA / NG, ARG / SA, ARG / NG and SA / NG. The method of maximum likelihood was used. The maximization of the log-likelihood functions was performed using the routine optim in the R software package (R Development Core Team [46]).

The folded and the corresponding truncated unfolded distributions have the same number of parameters. So, criteria like the Akaike information criterion and the Bayesian information criterion reduce to comparing log-likelihood values. In other words, the one giving the larger log-likelihood value can be regarded as the better model. Boxplots of the differences between the log-likelihood values for the forty five pairs are shown in Figure 1. We see that the differences are huge. They range from: 575.7203 to 865.9098 when the truncated bivariate normal and folded bivariate normal distributions are compared; 436.8568 to 744.1456 when the truncated bivariate t and folded bivariate t distributions are compared; 380.7097 to 865.9098 when the truncated bivariate skew normal and folded bivariate skew normal distributions are compared; 180.1524 to 299.9072 when the truncated bivariate skew t and folded bivariate skew t distributions are compared; 1275.391 to 2159.847 when the truncated bivariate logistic and folded bivariate logistic distributions are compared; 1243.677 to 2298.231 when the truncated bivariate Laplace and folded bivariate Laplace distributions of the first kind are compared; 401.2748 to 644.6660 when the truncated bivariate Kotz type and folded bivariate Kotz type distributions are compared; 4229.656 to 7849.704 when the truncated bivariate Laplace and folded bivariate Laplace distributions of the second kind are compared. This is compelling evidence that the folded distributions are much better models.

The fit of the eight folded distributions were compared in terms of log-likelihood values as well as the Akaike information criterion due to Akaike [2], the Bayesian information criterion due to Schwarz [47], the consistent Akaike information criterion (CAIC) due to Bozdogan [7], the corrected Akaike information criterion (AICc) due to Hurvich and Tsai [30],

the Hannan-Quinn criterion due to Hannan and Quinn [29] and p-values of the chisquared goodness of fit statistic.

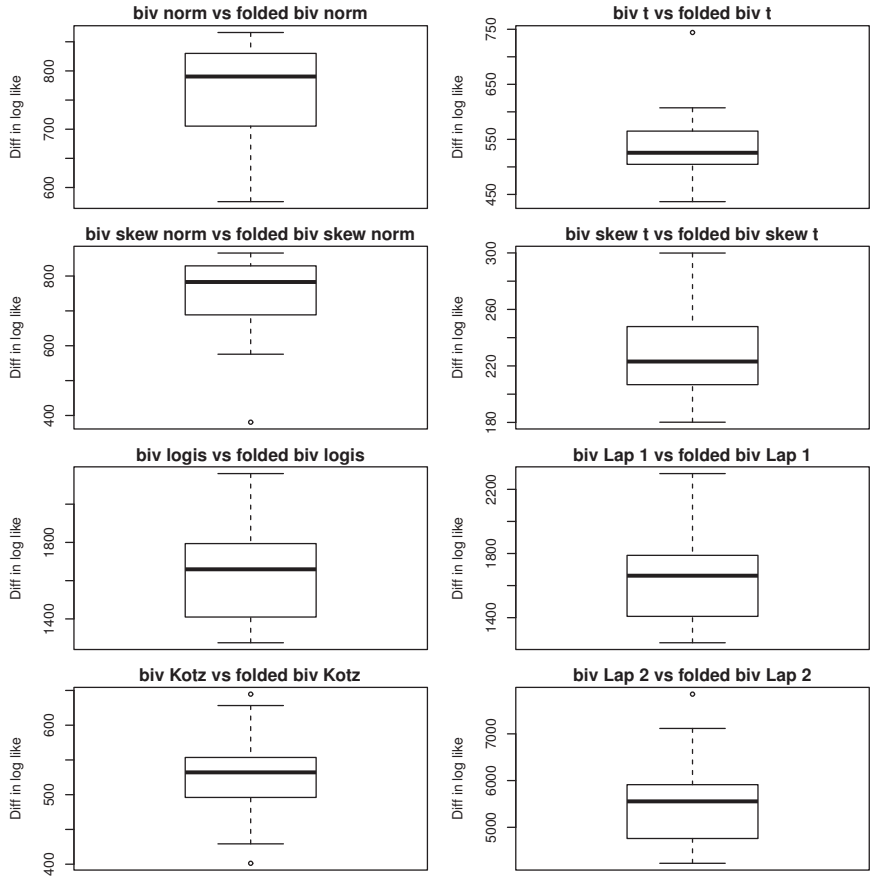


Figure 1: Boxplots of the differences between the log-likelihood values under the folded and truncated unfolded distributions for the forty five pairs.

The log-returns for each country exhibit heavy tails, presence of heavy tails was tested using the methods in Koning and Peng [35]. The p-values not reported here showed no evidence against heavy tails for each country. Hence, the log-returns for each pair of countries should be expected to exhibit heavy tails too. Of the eight distributions in Section 2, only the folded bivariate t and folded bivariate skew t distributions exhibit heavy tails.

For the following pairs (USA, CAD), (USA, ARG), (USA, SA), (USA, NG), (CAD, UK), (CAD, GER), (CAD, CHI), (CAD, JPN), (UK, GER), (UK, BRA), (UK, ARG), (GER, CHI), (GER, ARG), (CHI, JPN), (CHI, SA), (CHI, NG), (JPN, BRA), (JPN, ARG), (JPN, SA), (BRA, NG), (ARG, SA) and (SA, NG), the folded bivariate t distribution gave the best fit. The bivariate t distribution is heavy tailed but is also symmetric. The tests for bivariate symmetry (Snijders [48]) of the log-returns for these pairs showed no evidence against symmetry, see Table 3. This explains why the bivariate skew t distribution, a heavy tailed distribution accommodating for asymmetry, did not provide better fits for these pairs. None of the other distributions gave a significant p-value at the five percent significance level for each of these pairs.

For the following pairs (USA, UK), (USA, GER), (USA, CHI), (USA, JPN), (USA, BRA), (CAD, BRA), (CAD, ARG), (CAD, NG), (UK, SA), (GER, JPN), (GER, BRA), (GER, SA), (GER, NG), (CHI, ARG), (JPN, NG), (BRA, ARG), (BRA, SA) and (ARG, NG), the folded bivariate skew t distribution gave the best fit. The tests for bivariate symmetry of the log-returns for these pairs showed evidence against symmetry, see Table 3. This explains why the bivariate t distribution, a heavy tailed but symmetric distribution, did not provide better fits for these pairs. Again none of the other distributions gave a significant p-value at the five percent significance level for each of these pairs.

For the five remaining pairs of countries neither of the two heavy tailed distributions gave the best fit. For the pairs (CAD, SA), (UK, CHI), (UK, JPN) and (CHI, BRA), the folded bivariate Laplace distribution of the first kind gave the best fit. For the pair (UK, NG), the folded bivariate Kotz type distribution gave the best fit. The bivariate Laplace and Kotz type distributions are light tailed and are symmetric. We have not been to explain why these pairs were not best fitted by a heavy tailed distribution. However, the folded bivariate t distribution gave the second smallest values for AIC, BIC, CAIC, AICc, HQC and the second largest p-value for each of these pairs. Furthermore, the tests for bivariate symmetry for these pairs did not show evidence against symmetry, see Table 3.

Because of space concerns and to avoid repetitive discussion, we give details for only one of the forty five pairs, (USA, CAD). Table 1 gives the parameter estimates and standard errors for the fit of the eight folded distributions. Table 2 gives the log-likelihood values, AIC values, BIC values, CAIC values, AICc values, HQC values and p-values of the chisquared goodness of fit statistic for the fit of the eight folded distributions. We see that the folded bivariate t distribution gives the smallest values for AIC, BIC, CAIC, AICc and HQC and the largest p-value. The folded bivariate skew t distribution gives the second smallest values for AIC, BIC, CAIC, AICc and HQC. Contours of the joint probability density function of the best fitting distribution are shown in Figure 2. Also shown in this figure are the actual data values. The fit appears reasonable.

The magnitude correlations based on the best fits for the forty five pairs are given in Table 3. Also given in the table are p-values of the likelihood ratio test of the hypothesis that the absolute values of the components in each pair are independent. All of the correlations appear positive. This is expected since global economies are so inter dependent these days. One would not expect large stock values in magnitude for one country to be associated with small stock values in magnitude for another country or small stock values in magnitude for one country to be associated with large stock values in magnitude for another country. We also see that all of the correlations are significant except for (BRA, NG) and (ARG, NG). The strongest positive and significant correlations are for (UK, GER), (UK, SA) and (USA, CAD). The weakest positive and significant correlations are for (USA, NG), (CAD, NG), (UK, NG), (GER, NG), (CHI, NG), (JPN, NG) and (SA, NG).

We now give predictions based on bivariate value at risk curves. Under the folded t distribution, a bivariate value at risk curve with probability p is the solution of

$$\int_0^x \int_0^y f(u,v)dvdu = p, \tag{3.1}$$

where $f(u,v)$ is given by (2.3) with $(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\nu)$ replaced by $(\widehat{\mu_1},\widehat{\mu_2},\widehat{\sigma_1},\widehat{\sigma_2},\widehat{\rho},\widehat{\nu})$, the maximum likelihood estimates of $(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho,\nu)$. So, (x,y) satisfying (3.1) can be interpreted as the positive log returns for (USA, CAD) occurring with probability p. The curves of (x,y) are plotted in Figure 3 for $p=0.9,0.95,0.99$.

Table 1: Fitted models, parameter estimates and standard errors for USA / CAD.

Model	Parameter estimates (ses)
folded biv norm	$\widehat{\mu_1}=2.598\times10^{-5}(1.102\times10^{-2})$, $\widehat{\mu_2}=1.215\times10^{-4}(6.726\times10^{-3})$, $\widehat{s_1}=1.374\times10^{-2}(2.724\times10^{-4})$, $\widehat{s_2}=1.410\times10^{-2}(2.844\times10^{-4})$, $\widehat{\rho}=7.102\times10^{-1}(1.630\times10^{-2})$
folded biv t	$\widehat{\mu_1}=2.112\times10^{-3}(1.037\times10^{-3})$, $\widehat{\mu_2}=5.417\times10^{-3}(6.712\times10^{-4})$, $\widehat{s_1}=8.997\times10^{-3}(4.028\times10^{-4})$, $\widehat{s_2}=8.852\times10^{-3}(5.069\times10^{-4})$, $\widehat{\rho}=6.842\times10^{-1}(3.804\times10^{-2})$, $\widehat{\nu}=3.589(3.350\times10^{-1})$
folded biv skew norm	$\widehat{\mu_1}=2.599\times10^{-5}(1.437\times10^{-2})$, $\widehat{\mu_2}=1.215\times10^{-4}(2.831\times10^{-3})$, $\widehat{s_1}=1.374\times10^{-2}(2.724\times10^{-4})$, $\widehat{s_2}=1.410\times10^{-2}(3.447\times10^{-4})$, $\widehat{\rho}=7.102\times10^{-1}(1.989\times10^{-2})$, $\widehat{\alpha_1}=1.789\times10^{-11}(3.257)$, $\widehat{\alpha_2}=-1.953\times10^{-12}(1.823\times10^{-1})$
folded biv skew t	$\widehat{\mu_1}=2.112\times10^{-3}(1.354\times10^{-3})$, $\widehat{\mu_2}=5.417\times10^{-3}(8.358\times10^{-4})$, $\widehat{s_1}=8.997\times10^{-3}(3.266\times10^{-4})$, $\widehat{s_2}=8.852\times10^{-3}(2.341\times10^{-4})$, $\widehat{\rho}=6.842\times10^{-1}(3.144\times10^{-2})$, $\widehat{\alpha_1}=0.000(6.279\times10^{-1})$, $\widehat{\alpha_2}=0.000(6.301\times10^{-1})$, $\widehat{\nu}=3.589(3.283\times10^{-1})$
folded biv logis	$\widehat{\mu_1}=-3.405\times10^{-3}(2.427)$, $\widehat{\mu_2}=2.300\times10^{-3}(2.863)$, $\widehat{s_1}=60.4(4962.5)$, $\widehat{s_2}=72.3(6761.5)$
folded biv Lap 1	$\widehat{\mu_1}=13.4(8.4)$, $\widehat{\mu_2}=-13.0(2.3)$, $\widehat{s_1}=30.8(55.1)$, $\widehat{s_2}=34.7(77.3)$, $\widehat{\rho}=9.999\times10^{-1}(3.204\times10^{-2})$, $\widehat{r}=3.870\times10^{-3}(3.202\times10^{-6})$
folded biv Kotz	$\widehat{\mu_1}=4.678\times10^{-3}(4.393\times10^{-3})$, $\widehat{\mu_2}=6.530\times10^{-3}(8.504\times10^{-3})$, $\widehat{s_1}=9.928\times10^{-3}(2.001\times10^{-2})$, $\widehat{s_2}=9.500\times10^{-3}(1.003\times10^{-3})$, $\widehat{\rho}=5.823\times10^{-1}(5.332\times10^{-2})$, $\widehat{r}=200.3(6.2)$, $\widehat{s}=5.821\times10^{-2}(11.2)$, $\widehat{N}=6.838(4.522)$
folded biv Lap 2	$\widehat{\mu_1}=1.726\times10^{-4}(1.912\times10^{-3})$, $\widehat{\mu_2}=-2.919\times10^{-3}(2.455\times10^{-3})$, $\widehat{s_1}=9.928\times10^{-3}(7.655\times10^{-2})$, $\widehat{s_2}=9.500\times10^{-3}(5.366\times10^{-3})$, $\widehat{\rho}=9.999\times10^{-1}(9.022\times10^{-2})$

Table 2: Fitted models, log-likelihood values and selection criteria for USA / CAD.

Model	$-\ln L$	AIC	BIC	CAIC	AICc	HQC	p-value
folded biv norm	−10210.2	−20410.4	−20384.1	−20379.1	−20410.3	−20400.6	0.04
folded biv t	−10418.3	−20824.6	−20793.1	−20787.1	−20824.5	−20812.8	0.23
folded biv skew norm	−10210.2	−20406.4	−20369.6	−20362.6	−20406.3	−20392.7	0.05
folded biv skew t	−10418.3	−20820.6	−20778.6	−20770.6	−20820.5	−20804.9	0.23
folded biv logis	1710.3	3428.5	3449.5	3453.5	3428.5	3436.4	0.01
folded biv Lap 1	−6313.4	−12614.9	−12583.4	−12577.4	−12614.8	−12603.1	0.02
folded biv Kotz	−10406.9	−20797.8	−20755.8	−20747.8	−20797.7	−20782.1	0.04
folded biv Lap 2	−4518.0	−9026.0	−8999.7	−8994.7	−9025.9	−9016.2	0.03

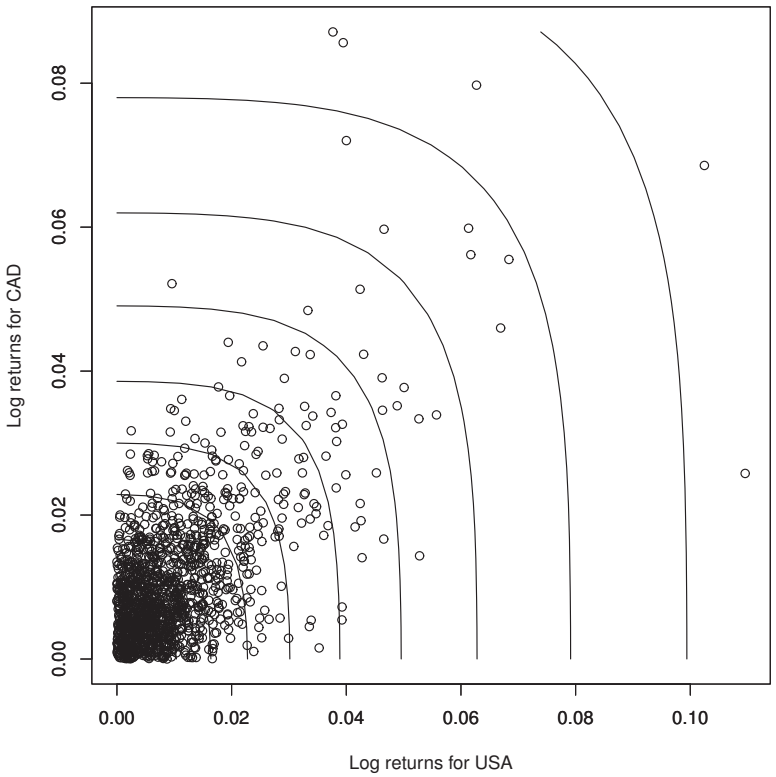


Figure 2: Contours of the joint probability density function of the fitted folded t distribution for (USA, CAD).

Finally, we check robustness of the fitted models by splitting the data into two halves. The first half was taken to be the data from 3rd January 2000 to 31 December 2007. The second half was taken to be the data from 1st January 2008 to 28th February 2014. We fitted the same models to each half. The results turned out to be the same as before. The folded bivariate t distribution gave the best fit for (USA, CAD), (USA, ARG), (USA, SA), (USA, NG), (CAD, UK), (CAD, GER), (CAD, CHI), (CAD, JPN), (UK, GER), (UK, BRA), (UK, ARG), (GER, CHI), (GER, ARG), (CHI, JPN), (CHI, SA), (CHI, NG), (JPN, BRA), (JPN, ARG), (JPN, SA), (BRA, NG), (ARG, SA) and (SA, NG) for each half. The folded bivariate skew t distribution gave the best fit for (USA, UK), (USA, GER), (USA, CHI), (USA, JPN), (USA, BRA), (CAD, BRA), (CAD, ARG), (CAD, NG), (UK, SA), (GER, JPN), (GER, BRA), (GER, SA), (GER, NG), (CHI, ARG), (JPN, NG), (BRA, ARG), (BRA, SA) and (ARG, NG) for each half. The folded bivariate Laplace distribution of the first kind gave the best fit for (CAD, SA), (UK, CHI), (UK, JPN) and (CHI, BRA) for each half. The folded bivariate Kotz type distribution gave the best fit for (UK, NG) for each half. The explanations for these best fits are the same as before.

Table 3: Estimated magnitude correlations, test for independence and test for bivariate symmetry.

Pair	Magnitude correlation	p-value for independence	p-value for symmetry
(USA, CAD)	0.596	0.000	0.112
(USA, UK)	0.499	0.000	0.006
(USA, GER)	0.533	0.000	0.048
(USA, CHI)	0.298	0.000	0.049
(USA, JPN)	0.153	0.000	0.001
(USA, BRA)	0.531	0.000	0.034
(USA, ARG)	0.317	0.000	0.086
(USA, SA)	0.366	0.000	0.057
(USA, NG)	0.049	0.003	0.074
(CAD, UK)	0.575	0.000	0.065
(CAD, GER)	0.509	0.000	0.095
(CAD, CHI)	0.342	0.000	0.077
(CAD, JPN)	0.189	0.000	0.084
(CAD, BRA)	0.507	0.000	0.021
(CAD, ARG)	0.312	0.000	0.013
(CAD, SA)	0.506	0.000	0.090
(CAD, NG)	0.057	0.001	0.004
(UK, GER)	0.743	0.000	0.068
(UK, CHI)	0.380	0.000	0.082
(UK, JPN)	0.210	0.000	0.062
(UK, BRA)	0.501	0.000	0.067
(UK, ARG)	0.268	0.000	0.081
(UK, SA)	0.626	0.000	0.042
(UK, NG)	0.051	0.002	0.066
(GER, CHI)	0.316	0.000	0.060
(GER, JPN)	0.175	0.000	0.006
(GER, BRA)	0.479	0.000	0.059
(GER, ARG)	0.259	0.000	0.055
(GER, SA)	0.529	0.000	0.008
(GER, NG)	0.043	0.009	0.002
(CHI, JPN)	0.370	0.000	0.075
(CHI, BRA)	0.306	0.000	0.076
(CHI, ARG)	0.152	0.000	0.034
(CHI, SA)	0.372	0.000	0.058
(CHI, NG)	0.064	0.000	0.091
(JPN, BRA)	0.145	0.000	0.094
(JPN, ARG)	0.097	0.000	0.056
(JPN, SA)	0.236	0.000	0.075
(JPN, NG)	0.075	0.000	0.009
(BRA, ARG)	0.351	0.000	0.028
(BRA, SA)	0.435	0.000	0.019
(BRA, NG)	0.022	0.191	0.077
(ARG, SA)	0.215	0.000	0.083
(ARG, NG)	0.020	0.222	0.012
(SA, NG)	0.041	0.012	0.067

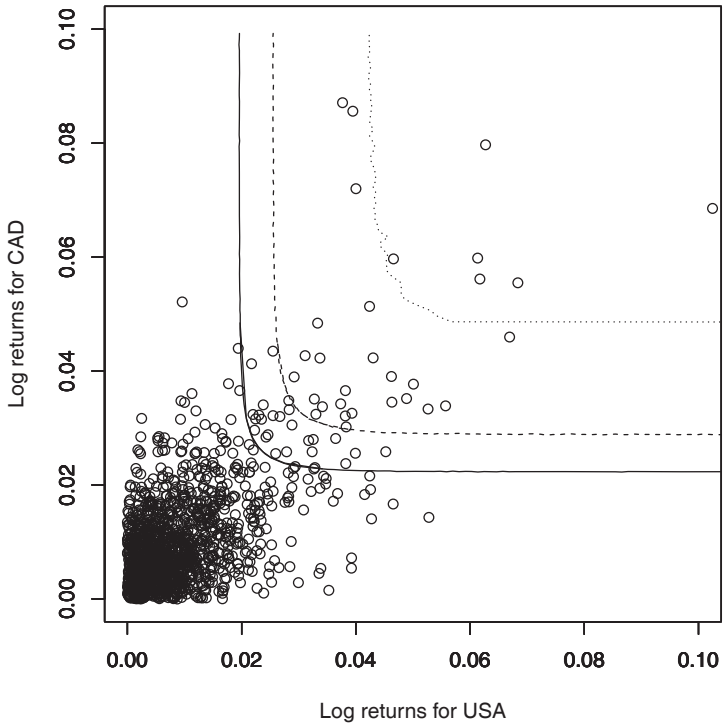


Figure 3: Value at risk curves of the fitted folded t distribution at $p = 0.9, 0.95, 0.99$ for (USA, CAD).

4. CONCLUSIONS

Motivated by the concept of magnitude correlation of stock returns, we have introduced the following folded bivariate distributions: the folded bivariate skew normal distribution; the folded bivariate skew t distribution; the folded bivariate logistic distribution; the folded bivariate Kotz type distribution; the folded bivariate Laplace distribution of the first kind; the folded bivariate Laplace distribution of the second kind. We have also introduced the following folded univariate distributions: the folded univariate skew normal distribution; the folded univariate skew t distribution.

We fitted eight folded bivariate distributions to forty five real data sets. The two heavy tailed distributions, the folded bivariate t and folded bivariate skew t distributions, gave the best fit for forty of the data sets. The remaining five data sets were best fitted by folded bivariate Laplace distribution of the first kind and the folded bivariate Kotz type distribution, two of the lighted tailed distributions. We have not been able to explain why these five data sets were best fitted by light tailed distributions when all of the data sets are heavy tailed.

We also compared the fits of the folded and truncated unfolded distributions using the same data sets. Remarkably each folded distribution outperformed the corresponding truncated unfolded distribution for each of the forty five data sets. This shows that magnitude correlations can be better modeled by folded bivariate distributions.

A future work is to extend the results of this paper for folded multivariate distributions, folded matrix variate distributions and folded complex variate distributions.

ACKNOWLEDGMENTS

The authors would like to thank the Editor, the Associate Editor and the referee for careful reading and comments which greatly improved the paper.

REFERENCES

[1] Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, DC.

[2] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 716–723.

[3] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society B*, **65**, 367–389.

[4] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715–726.

[5] Balakrishnan, N. and Lai, C.D. (2009). *Continuous Bivariate Distributions*, Second Edition, Springer Verlag, New York.

[6] Bhamra, H.S.; Coeurdacier, N. and Guibaud, S. (2014). A dynamic equilibrium model of imperfectly integrated financial markets, *Journal of Economic Theory*, **154**, 490–542.

[7] Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions, *Psychometrika*, **52**, 345–370.

[8] Breusch, T.S. (1979). Testing for autocorrelation in dynamic linear models, *Australian Economic Papers*, **17**, 334–355.

[9] Breusch, T.S. and Pagan, A.R. (1979). A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, **47**, 1287–1294.

[10] Burnham, K.P. and Anderson, D.R. (2004). Multimodel inference: understanding AIC and BIC in model selection, *Sociological Methods and Research*, **33**, 261–304.

[11] Chabakauri, G. (2013). Dynamic equilibrium with two stocks, heterogeneous investors, and portfolio constraints, *Review of Financial Studies*, **26**, 3104–3141.

[12] Chakraborty, A.K. and Chatterjee, M. (2013). On multivariate folded normal distribution, *Sankhyā B*, **75**, 1–15.

[13] Cooray, K.; Gunasekera, S. and Ananda, M.M.A. (2006). The folded logistic distribution, *Communications in Statistics – Theory and Methods*, **35**, 385–393.

[14] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

[15] Cox, D.R. and Stuart, A. (1955). Some quick sign test for trend in location and dispersion, *Biometrika*, **42**, 80–95.

[16] Durbin, J. and Watson, G.S. (1950). Testing for serial correlation in least squares regression I, *Biometrika*, **37**, 409–428.

[17] Durbin, J. and Watson, G.S. (1951). Testing for serial correlation in least squares regression II, *Biometrika*, **38**, 159–178.

[18] Durbin, J. and Watson, G.S. (1971). Testing for serial correlation in least squares regression III, *Biometrika*, **58**, 1–19.

[19] Eltoft, T.; Kim, T. and Lee, T.-W. (2006). On the multivariate Laplace distribution, *IEEE Signal Processing Letters*, **13**, 300–303.

[20] Ernst, M.D. (1998). A multivariate generalized Laplace distribution, *Computational Statistics*, **13**, 227–232.

[21] Fang, Y. (2011). Asymptotic equivalence between cross-validations and Akaike Information Criteria in mixed-effects models, *Journal of Data Science*, **9**, 15–21.

[22] Firth, M. (1996). Dividend changes, abnormal returns, and intra-industry firm valuations, *The Journal of Financial and Quantitative Analysis*, **31**, 189–211.

[23] Fukuda, S. and Tanaka, M. (2013). Financial crises and risk premiums in international interbank markets, *Policy Research Institute, Ministry of Finance, Japan, Public Policy Review*, **9**, 117–138.

[24] Godfrey, L.G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica*, **46**, 1293–1302.

[25] Gu, J.; Jones, J.S. and Liu, P. (2014). Do credit rating agencies sacrifice timeliness by pursuing rating stability? Evidence from equity market reactions to CreditWatch events, *Theoretical Economics Letters*, **4**, 311–322.

[26] Guillou, A.; Naveau, P. and You, A. (2015). A folding methodology for multivariate extremes: estimation of the spectral probability measure and actuarial applications, *Scandinavian Actuarial Journal*, **7**, 549–572.

[27] Gumbel, E.J. (1961). Bivariate logistic distributions, *Journal of the American Statistical Association*, **56**, 335–349.

[28] Hamalainen, J. (2015). Correlation in the magnitude of financial returns, University of Turku, Finland.

[29] Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society B*, **41**, 190–195.

[30] Hurvich, C.M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples, *Biometrika*, **76**, 297–307.

[31] Inman, H.F. and Bradley, E.L. (1991). Approximations to the mean and variance of the index of dissimilarity in $2 \times C$ tables under a random allocation model, *Sociological Methods Research*, **20**, 242–255.

[32] Islam, M.; Ali, L.E. and Afroz, N. (2012). Forecasting volatility of Dhaka stock exchange: linear vs non-linear models, *International Journal of Science and Engineering*, **3**, 4–8.

[33] Jamalizadeh, A.; Khosravi, M. and Balakrishnan, N. (2009). Recurrence relations for distributions of a skew-t and a linear combination of order statistics from a bivariate-t, *Computational Statistics and Data Analysis*, **53**, 847–852.

[34] Jamalizadeh, A.; Mehrali, Y. and Balakrishnan, N. (2009). Recurrence relations for bivariate t and extended skew-t distributions and an application to order statistics from bivariate t, *Computational Statistics and Data Analysis*, **53**, 4018–4027.

[35] Koning, J. and Peng, L. (2008). Goodness-of-fit tests for a heavy tailed distribution, *Journal of Statistical Planning and Inference*, **138**, 3960–3981.

[36] Kotz, S. (1975). *Multivariate distributions at a cross-road*. In: "Statistical Distributions in Scientific Work" (G.P. Patil, S. Kotz and J.K. Ord, Eds.), Vol. 1, D. Reidel Publishing Company, Dordrecht, The Netherlands.

[37] Kotz, S. and Ostrovskii, I. (1994). Characteristic functions of a class of elliptical distributions, *Journal of Multivariate Analysis*, **49**, 164–178.

[38] Kourtis, A. (2014). On the distribution and estimation of trading costs, *Journal of Empirical Finance*, **28**, 104–117.

[39] Kutergin, A. and Filimonov, V. (2015). *On the modeling of financial time series*. In: "Financial Econometrics and Empirical Market Microstructure" (A.K. Bera *et al.*, Eds.), pp. 131–151.

[40] Leone, F.C.; Nelson, L.S. and Nottingham, R.B. (1961). The folded normal distribution, *Technometrics*, **3**, 543–550.

[41] Mulekar, M.S.; Knutson, J.C. and Champanerkar, J.A. (2008). How useful are approximations to mean and variance of the index of dissimilarity? *Computational Statistics and Data Analysis*, **52**, 2098–2109.

[42] Muralidhar, G.S.; Bovik, A.C. and Markey, M.K. (2013). Noise analysis of a new singularity index, *IEEE Transactions on Signal Processing*, **61**, 6150–6163.

[43] Nadarajah, S. and Bakar, S.A.A. (2015). New folded models for the log-transformed Norwegian fire claim data, *Communications in Statistics – Theory and Methods*, **44**, 4408–4440.

[44] Psarakis, S. and Panaretos, J. (1990). The folded t distribution, *Communications in Statistics – Theory and Methods*, **19**, 2717–2734.

[45] Psarakis, S. and Panaretos, J. (2001). On some bivariate extensions of the folded normal and the folded-t distributions, *Journal of Applied Statistical Science*, **10**, 119–136.

[46] R Development Core Team (2020). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[47] Schwarz, G.E. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–464.

[48] Snijders, T. (1981). Rank tests for bivariate symmetry, *Annals of Statistics*, **9**, 1087–1095.

[49] Tsai, K.T.; Lih, J.S. and Ko, J.Y. (2012). The overnight effect on the Taiwan stock market, *Physica A – Statistical Mechanics and Its Applications*, **391**, 6497–6505.

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 39–62
https://doi.org/10.57805/revstat.v21i1.394

Estimation of Distribution Function Using Percentile Ranked Set Sampling

Authors: Yusuf Can Sevil ✉
– The Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey
yusuf.sevil@ogr.deu.edu.tr

Tugba Ozkal Yildiz
– Department of Statistics, Dokuz Eylul University, Izmir, Turkey
tugba.ozkal@deu.edu.tr

Received: December 2020 Revised: July 2021 Accepted: July 2021

Abstract:

- The estimation of distribution function has received considerable attention in the literature. Because, many practical problems involve estimation of distribution function from experimental data. Estimating the distribution function makes it possible to do pointwise estimation and to make statistical inference about the distribution of interested population. In this study, we suggested an empirical distribution function (EDF) for percentile ranked set sampling (PRSS). Bias of the EDF estimator is investigated theoretically and numerically. Relative efficiencies of the proposed EDF estimator based on PRSS with respect to EDF estimator based on simple random sampling (SRS) and ranked set sampling (RSS) are obtained. We also considered impact of imperfect rankings on the EDF based on PRSS. According to the results, the proposed EDF estimator is unbiased for the extreme "minimum and maximum" points and center of the distribution. Also, it is clearly appeared that the EDF estimator based on PRSS is more efficient than the EDF based on SRS. Another important result is that the suggested EDF estimator has larger efficiencies than the EDF based on RSS for some special cases of PRSS. In the application, the EDF based on PRSS is used to estimate the proportion of women in obesity class III (BMI$>$ 40).

Keywords:

- *percentile ranked set sampling; empirical distribution function; relative efficiency; mean squared error; imperfect ranking; body mass index data.*

AMS Subject Classification:

- 62P10, 62D99, 68U20.

✉ Corresponding author.

1. INTRODUCTION

Ranked set sampling (RSS) was introduced by McIntyre [13] as an advantageous alternative to simple random sampling (SRS). McIntyre [13] studied mean estimator based on RSS and showed that this estimator is more efficient than mean estimator using SRS. Then, mathematical theory of RSS was first suggested by Takahasi and Wakimoto [25]. By Dell and Clutter [6], it was proved that mean estimator based on RSS is unbiased and more efficient than mean estimator based on SRS even if ranking is not perfect. In the literature, there are some other estimators based on RSS such as estimation of correlation coefficient [22], estimation of variance [23] and estimation of population proportion [15, 29, 30]. Also, for more extended literature about RSS, see Kaur *et al.* [11] and Al-Omari and Bouza [3].

The estimation of cumulative distribution function (CDF) with various settings of the RSS has been studied by many authors. Stokes and Sager [24] suggested an unbiased estimator based on RSS for population distribution function. Samawi and Al-Sagheer [19] considered EDF estimator based on extreme ranked set sampling and median ranked set sampling. EDF using double ranked set sampling was investigated by Abu-Dayyeh *et al.* [1]. Al-Omari [2] studied EDF based on quartile ranked set sampling. Sevil and Yildiz [20] developed estimation of distribution function using RSS based on level-2 sampling design. Also, Kolmogorov Smirnov (KS) test statistic based on RSS was compared with KS test statistic based on SRS by Sevil and Yildiz [20]. EDF estimators using RSS based on three different sampling designs were given by Yildiz and Sevil [26, 27]. Some goodness of fit tests based on these EDF estimators were investigated in their study. Sevil and Yildiz [21] discussed design-based estimators based on level-0, level-1 and level-2 for distribution function of finite population. Some other distribution function estimators were considered for extreme median ranked set sampling [12], selective order ranked set sampling [4], partially rank-ordered set [16] and pair ranked set sampling [28].

By using percentiles instead of quartiles, more flexible selection procedure named as percentile ranked set sampling (PRSS) was suggested by Muttlak [14]. In Muttlak's study, estimation of mean is investigated using PRSS. Since PRSS is general form of quartile ranked set sampling (QRSS) and extreme ranked set sampling (ERSS), EDF estimators based on QRSS and ERSS can be obtained by using EDF estimator based on PRSS. Moreover, EDF estimator based on median ranked set sampling (MRSS) can be derived by using EDF estimator of PRSS when the set size is even. So, the EDF estimator using PRSS becomes quite useful estimator. Therefore, we considered the performance of EDF estimator using PRSS under perfect and imperfect rankings.

This study is organized as follows. In section two, PRSS procedure is defined. The EDF estimator based on PRSS is given in section three. Also, the properties of the EDF estimator are discussed. In section four, we introduce Frey's one-parameter ranking error model [7] and study imperfect ranking case for proposed EDF estimator. Also, we obtained some results under imperfect ranking in this section. Some inferences about CDF, $F(x)$, are given in section five. Moreover, body mass index data is used to illustrate the EDF using PRSS. Finally, some conclusion remarks are stated in section six.

2. PERCENTILE RANKED SET SAMPLING

Muttlak [14] proposed PRSS as practical sampling scheme according to RSS. In literature, modified versions of PRSS can be seen such as double PRSS [9] and multistage PRSS [10].

In this method, p-th and q-th percentile of the sample are selected for full measurement, $0 < p < 1$ and $q = 1 - p$. Before we describe the procedure of PRSS, we give some notations. Let k, l and n denote set size, number of cycles and total sample size, respectively. Also, $(X_{11j}, X_{12j}, ..., X_{1kj})$, $(X_{21j}, X_{22j}, ..., X_{2kj})$, ..., $(X_{k1j}, X_{k2j}, ..., X_{kkj})$ are random sets of size k from j-th cycle, $j = 1, ..., l$. Here, it is assumed that X_{itj} is selected from a population with continuous density function $f(x)$ and CDF $F(x)$. The order statistics of the i-th set are denoted by $X_{i(1)j}, X_{i(2)j}, ..., X_{i(k)j}$, $i = 1, ..., k$.

Now, we define the procedure of PRSS. First, k^2 units are selected without replacement from the population. These units are divided into the k random sets, each of size k. In each set, these units are ranked from the smallest to the largest. If the set size k is odd, PRSS is denoted by $PRSS_O$ and it is obtained by using the following steps:

- (**i**) From the first $(k-1)/2$ sets, the r-th smallest units are measured, $X_{(r)}$;
- (**ii**) The median ranked unit is measured from the $((k+1)/2)$-th set, $X_{(m)}$;
- (**iii**) Then, the s-th smallest units are measured from the remaining $(k-1)/2$ sets, $X_{(s)}$;

where r and s are the nearest integer value of $p(k+1)$ and $q(k+1)$, respectively. Note that $r = 1$ if $p(k+1) < 0.5$ and $s = k$ if the nearest integer value of $q(k+1)$ is larger than k. If the set size k is even, PRSS is denoted by $PRSS_E$ and it is obtained by using the following steps.

- (**i**) From the first $k/2$ sets, the r-th smallest units are measured, $X_{(r)}$;
- (**ii**) Then, the s-th smallest units are measured from the remaining $k/2$ sets, $X_{(s)}$.

To obtain $n = lk$ sample observations, these procedures are repeated l times. $PRSS_O$ and $PRSS_E$ are denoted by

$$PRSS_O = \left\{X_{1(r)j},\ X_{2(r)j},\ ...,\ X_{\frac{k-1}{2}(r)j},\ X_{m(m)j},\ X_{\frac{k+3}{2}(s)j},\ ...,\ X_{k-1(s)j},\ X_{k(s)j}\right\}$$

and

$$PRSS_E = \left\{X_{1(r)j},\ ...,\ X_{\frac{k}{2}(r)j},\ X_{\frac{k+2}{2}(s)j},\ ...\ , X_{k(s)j}\right\},$$

respectively, where $m = (k+1)/2$ and $j = 1, ..., l$.

As defined in Stokes and Sager [24], lk independent copies (Y, R) are observed as follows: R is first selected at random from $1, ..., k$ and Y is observed according to $F_{(i)}(x)$ (the CDF of the i-th order statistics), then the marginal joint distribution of Y's is the same as that of the SRS. This statement is given in the part (a) of the Theorem 1 by Stokes and Sager [24]. Part (b) of the Theorem 1 capitalizes on this characterization to link RSS with SRS.

Let $T^{'}=(T_1,T_2,...,T_k)$ be a multinomial random vector with lk trials and $P=(\frac{1}{k},\frac{1}{k},...,\frac{1}{k})$ be a probability vector. It is supposed that the lk random variables were obtained by first observing T and then selecting T_i units randomly from a population with probability density function (PDF) $f_{(i)}(x)$, $i=1,...,k$. Also, the obtained lk units are denoted by $Y_1,Y_2,...,Y_{lk}$.

Theorem 2.1. *With the same conditions of Theorem 1 in Stokes and Sager [24], we give the following:*

(**1**) *When the set size is odd,* $\{Y_1,Y_2,...,Y_{lk}\,|\,T=(0,...,0,t_r=\frac{(k-1)l}{2},0,...,0,t_m=l,0,...,0,t_s=\frac{(k-1)l}{2},0,...,0)\}$ *has the same probability structure as* $\{X_{g(r)j},X_{m(m)j},X_{h(s)j};g=1,2,...,\frac{k-1}{2};m=\frac{k+1}{2};h=\frac{k+3}{2},\frac{k+5}{2},...,k;j=1,2,...,l\}$ *where ranks of the measured observations could be one of the* (r,s) *pairs,* $\{(1,k),(2,k-1),...,(\frac{k-1}{2},\frac{k+3}{2})\}$.

(**2**) *When set size is even,* $\{Y_1,Y_2,...,Y_{lk}\,|\,T=(0,...,0,t_r=\frac{lk}{2},0,...,0,t_s=\frac{lk}{2},0,...,0)\}$ *has the same probability structure as* $\{X_{g(r)j},X_{h(s)j};g=1,2,...,\frac{k}{2};h=\frac{k+2}{2},\frac{k+4}{2},...,k;j=1,2,...,l\}$ *where ranks of the measured observations could be one of the* (r,s) *pairs,* $\{(1,k),(2,k-1),...,(\frac{k}{2},\frac{k+2}{2})\}$.

These parts (1) and (2) are proved in Appendices.

3. EMPIRICAL DISTRIBUTION FUNCTION OF PERCENTILE RANKED SET SAMPLING

In this section, we described the suggested EDF estimator based on PRSS. Also, properties of the EDF estimator are given. Bias and efficiency of the EDF based on PRSS are investigated and compared with distribution function estimators using SRS and RSS. It is assumed that $X_1,X_2,...,X_n$ be a simple random sample. EDF based on SRS is denoted by $\hat{F}_{SRS}(x)$,

$$\hat{F}_{SRS}(x)=\frac{1}{n}\sum_{i=1}^{n}I(X_i\le x).$$

where $I(.)$ is indicator function. The EDF based on SRS is unbiased estimator of $F(x)$ for given x, with variance $V(\hat{F}_{SRS}(x))=\frac{1}{n}F(x)(1-F(x))$.

Stokes and Sager [24] proposed $\hat{F}_{RSS}(x)$ for estimating the distribution function $F(x)$. Let $\{X_{1(1)j},X_{2(2)j},...,X_{k(k)j}\}$ be the order statistics that are obtained by using RSS,

$$\hat{F}_{RSS}(x)=\frac{1}{lk}\sum_{j=1}^{l}\sum_{i=1}^{k}I\big(X_{i(i)j}\le x\big) \tag{3.1}$$

They showed that $\hat{F}_{RSS}(x)$ is unbiased with variance

$$V\Big(\hat{F}_{RSS}(x)\Big)=\frac{1}{lk^2}\sum_{i=1}^{k}F_{(i)}(x)\big(1-F_{(i)}(x)\big)$$

where $F_{(i)}(x)$ is distribution function of the i-th order statistic, and

$$\frac{\hat{F}_{RSS}(x) - E\Big(\hat{F}_{RSS}(x)\Big)}{\Big(V\Big(\hat{F}_{RSS}(x)\Big)\Big)^{1/2}}$$

converges in distribution to standard normal as $l \to \infty$, when x and k are held fixed.

Let $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$ are the EDFs of a PRSS data when set size is odd and even, respectively. If set size is odd,

(3.2)
$$\hat{F}_{PRSS_O}(x) = \frac{1}{lk}\left[\sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} I\big(X_{i(r)j} \le x\big) + \sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} I\Big(X_{\frac{k+1}{2}+i(s)j} \le x\Big) + \sum_{j=1}^{l} I\big(X_{m(m)j} \le x\big)\right]$$

and if set size is even,

$$\hat{F}_{PRSS_E}(x) = \frac{1}{lk}\left[\sum_{j=1}^{l}\sum_{i=1}^{\frac{k}{2}} I\big(X_{i(r)j} \le x\big) + \sum_{j=1}^{l}\sum_{i=1}^{\frac{k}{2}} I\Big(X_{\frac{k}{2}+i(s)j} \le x\Big)\right] \tag{3.3}$$

where $r \approx p(k+1)$, $s \approx q(k+1)$ and $m = \frac{k+1}{2}$ is the median ranked unit. Under the perfect ranking, we state the following propositions for some basic properties of these distribution function estimators.

Proposition 3.1.

(**a**) *Using* $PRSS_O$

i. $E\Big(\hat{F}_{PRSS_O}(x)\Big) = \big(\frac{1}{2} - \frac{1}{2k}\big)\big(F_{(r)}(x) + F_{(s)}(x)\big) + \frac{1}{k}F_{(m)}(x)$,

ii. $V\Big(\hat{F}_{PRSS_O}(x)\Big) = \frac{1}{lk^2}\left[\begin{array}{c}\big(\frac{k-1}{2}\big)\big(F_{(r)}(x)\big(1-F_{(r)}(x)\big) + F_{(s)}(x)\big(1-F_{(s)}(x)\big)\big) + \\ F_{(m)}(x)\big(1-F_{(m)}(x)\big)\end{array}\right]$;

(**b**) *Using* $PRSS_E$

i. $E\Big(\hat{F}_{PRSS_E}(x)\Big) = \frac{1}{2}\big(F_{(r)}(x) + F_{(s)}(x)\big)$,

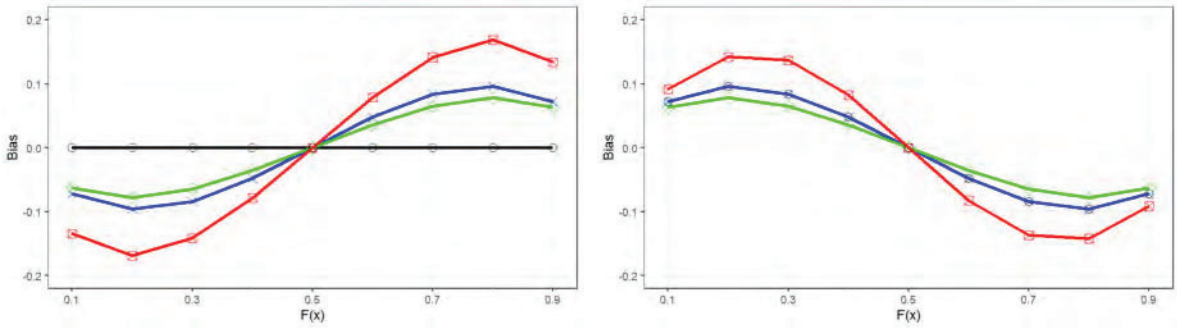
ii. $V\Big(\hat{F}_{PRSS_E}(x)\Big) = \frac{1}{2lk}\big[F_{(r)}(x)\big(1-F_{(r)}(x)\big) + F_{(s)}(x)\big(1-F_{(s)}(x)\big)\big]$;

where $F_{(r)}(x)$, $F_{(s)}(x)$ *and* $F_{(m)}(x)$ *are distribution function of* $X_{(r)}$, $X_{(s)}$ *and* $X_{(m)}$, *respectively.*

Part (a) and part (b) are proved in Appendices. As seen in Proposition 3.1, $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$ are biased estimators for $F(x)$. However, the bias is almost zero as $F(x)$ gets closer to 1, 0.5 and 0 under perfect ranking. Also, the biases of these estimators do not depend on the number of cycles. The biases of these EDFs can be calculated by using following equations.

$$Bias[\hat{F}_{PRSS_O}(x)] = F(x) - E(\hat{F}_{PRSS_O}(x)), \tag{3.4}$$

$$Bias[\hat{F}_{PRSS_E}(x)] = F(x) - E(\hat{F}_{PRSS_E}(x)). \tag{3.5}$$



(a) When $p = 0.1$, the bias of EDFs. **(b)** When $p = 0.4$, the bias of EDFs.

Figure 1: Bias for $\hat{F}_{PRSS_O}$ and $\hat{F}_{PRSS_E}$ where black, blue, green and red curves are $k = 3$, $k = 4$, $k = 5$ and $k = 6$, respectively.

These biases of $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$ are given by Figure 1 when $p = 0.1$ and $p = 0.4$. These EDF estimators are unbiased as $F(x)$ gets closer to 1, 0.5 and 0. The bias increases as k increases except for $F(x) = 0.5$. In the Figure 1(b), the blue and black curves are overlapping.

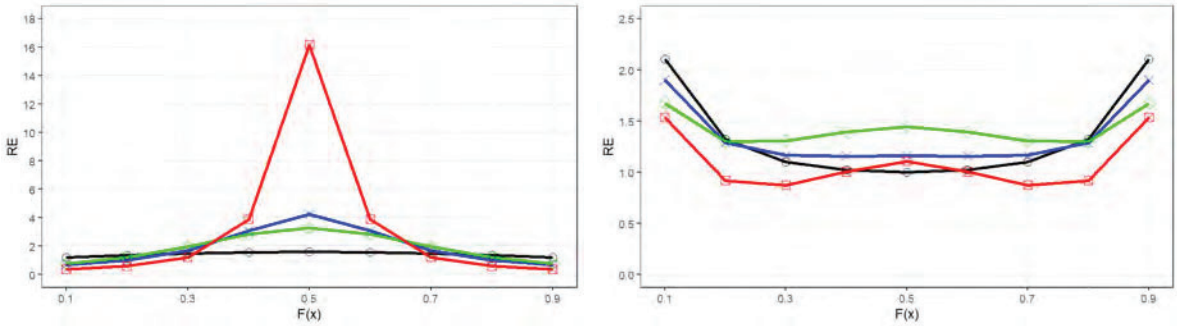
Mean squared error is used as a measure of performance of the proposed estimators. Then, relative efficiencies (RE) of $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$ with respect to $\hat{F}_{SRS}(x)$ are described as

$$RE[\hat{F}_{PRSS_O}(x), \hat{F}_{SRS}(x)] = \frac{V(\hat{F}_{SRS}(x))}{MSE(\hat{F}_{PRSS_O}(x))},$$

and

$$RE[\hat{F}_{PRSS_E}(x), \hat{F}_{SRS}(x)] = \frac{V(\hat{F}_{SRS}(x))}{MSE(\hat{F}_{PRSS_E}(x))}.$$

REs are illustrated by the Figure 2. When $p = 0.1$, it is seen that the REs peak on the middle of the distribution function. Even, the EDFs based on PRSS are more efficient than the EDF based on RSS whenever $F(x)$ is close to 0.5 comparing with Stokes and Sager [24]. The REs increase while the set size increases. When $p = 0.4$, Figure 2 shows that the REs are higher on the tails of the distribution function. Whenever $F(x)$ is close to 0.1 (or 0.9) comparing with Stokes and Sager [24], the EDFs based on PRSS are more efficient than the EDF based on RSS. Also, the REs are almost equal to or larger than 1 for any $F(x)$ when $k = 3, 4, 5, 6$ and $p = 0.4$.



(a) When $p = 0.1$, the REs of EDFs. **(b)** When $p = 0.4$, the REs of EDFs.

Figure 2: REs for $\hat{F}_{PRSS_O}$ and $\hat{F}_{PRSS_E}$ where black, blue, green and red curves are $k = 3$, $k = 4$, $k = 5$ and $k = 6$, respectively.

Table 1 indicates REs of EDFs using PRSS when $F(x) = 0.1$ and $F(x) = 0.5$ relative to RSS. The REs are obtained by using the following equations.

$$RE[\hat{F}_{PRSS_O}(x), \hat{F}_{RSS}(x)] = \frac{V(\hat{F}_{RSS}(x))}{MSE(\hat{F}_{PRSS_O}(x))},$$

and

$$RE[\hat{F}_{PRSS_E}(x), \hat{F}_{RSS}(x)] = \frac{V(\hat{F}_{RSS}(x))}{MSE(\hat{F}_{PRSS_E}(x))}.$$

It can be shown that the EDFs based on PRSS (with $p = 0.4$) have higher performances than the EDF based on RSS when $F(x) = 0.1$. Also, the EDFs using PRSS (with $p = 0.1$) are more efficient than the EDF using RSS when $F(x) = 0.5$.

Table 1: The REs of the EDF estimators based on PRSS with respect to RSS.

	$F(x) = 0.1$		$F(x) = 0.5$	
k	$p = 0.1$	$p = 0.4$	$p = 0.1$	$p = 0.4$
3	1.000	1.000	1.760	0.625
4	0.522	2.333	1.473	0.636
5	0.557	1.635	1.227	0.720
6	0.263	7.303	1.045	0.500

The following proposition is needed to study some asymtotic inference about the expected value of the estimators, $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$. The Proposition 3.2 is proved in Appendices.

Proposition 3.2. *For fixed k and $l \to \infty$, the following results are obtained:*

(**a**) $\frac{\hat{F}_{PRSS_O}(x) - E\left(\hat{F}_{PRSS_O}(x)\right)}{\sqrt{V\left(\hat{F}_{PRSS_O}(x)\right)}}$ *converges in distribution to* $N(0,1)$;

(**b**) $\frac{\hat{F}_{PRSS_E}(x) - E\left(\hat{F}_{PRSS_E}(x)\right)}{\sqrt{V\left(\hat{F}_{PRSS_E}(x)\right)}}$ *converges in distribution to* $N(0,1)$.

4. IMPERFECT RANKING

The efficiency of PRSS is affected by ranking steps. In general, the ranking is performed by subjective judgement or according to concomitant (auxiliary) variable that is correlated to the variable of interest. In the ranking steps, it is assumed that the ranking is completely accurate. However, this is not a realistic assumption. Therefore, one of the interesting topic is ranking error models in the literature. Dell and Clutter [6] proposed adaptive perceptual error model. Bohn and Wolfe [5] suggested ranking error model that constructs the judgement class distributions as a mixture distribution of the actual order statistics. Then, Frey [7, 8] extended the model [5] and introduced new class of models for imperfect ranking. Ozturk [17] estimated the parameters of ranking error models of Bohn and Wolfe [5] and Frey [7, 8].

He proved that one-parameter ranking error model [7, 8] is more efficient than ranking error model [5].

In this section, we investigated the effect of imperfect ranking on PRSS using Frey's one-parameter judgement ranking [7]. It is assumed that $k!$ possible judgment orderings of the true order statistics $X_{(i_1:h)}, ..., X_{(i_k:h)}$ selected from a larger set of size h, $h \geq k$. Random selection of set of size k yields $\binom{h}{k}$ possible selection of k order statistics out of h order statistics in the larger set and all these selections are equally likely. Let $\boldsymbol{A}(i_1, ..., i_k)$ be a doubly stochastic matrix. Frey [7] specified a way to compute the matrix $\boldsymbol{A}(i_1, ..., i_k)$,

$$\boldsymbol{A}(i_1, ..., i_k) = \frac{1}{k!} \sum_{\pi \in S_k} q(i_{\pi(1)}, ..., i_{\pi(k)}) \times Per(\pi(1), ..., \pi(k)),$$

where $q(i_{\pi(1)}, ..., i_{\pi(k)})$ denotes the probability that corresponds to the ordering of $X_{(i_1:h)} < \cdots < X_{(i_k:h)}$, $Per(\pi(1), ..., \pi(k))$ is the permutation matrix whose $(i, \pi(i))$-th entry is one for $i = 1, ..., k$ and all other entries are zero, and S_k is the set of all permutations. The probabilities $q(i_{\pi(1)}, ..., i_{\pi(k)})$ are obtained by selecting an appropriate weight function $w(\pi)$ with $\pi \in S_k$. These weights must be normalized, so these are actually probabilities. A class of weight function was suggested by Frey [7],

$$w(\pi) = \exp\left\{ \delta \sum_{j=1}^{k} j\lambda\left(\frac{i_{\pi(j)}}{h+1} \right) \right\}$$

where δ is called as power and $\delta \in [0, \infty)$. When $\delta = 0$, a completely random ranking model is constructed. When δ approaches infinity, the probability $q(i_{\pi(1)}, ..., i_{\pi(k)})$ concentrates on the single permutation having the largest value of

$$\sum_{j=1}^{k} j\lambda\left(\frac{i_{\pi(j)}}{h+1} \right)$$

and corresponds to a perfect ranking model. Also, a wide range of imperfect ranking models can be obtained using the other values of δ. Frey [7] proposed three different λ function which are $\lambda_1(u) = u$, $\lambda_2(u) = -u^{-1}$ and $\lambda_3(u) = (1-u)^{-1}$ to obtain symmetric, skewed-left and skewed-right imperfect ranking probabilities. Note that these probabilities do not depend on shape of underlying distributions. $\boldsymbol{\Omega}(i_1, ..., i_k)$ is a $k \times h$ matrix to exhaust the selection of all possible judgment orderings. In this matrix, the $(i', i_{i'})$-th entry is one for $i' = 1, ..., k$ and all other entries are zero. Then, the matrix product

$$\boldsymbol{N}(i_1, ..., i_k) = \boldsymbol{A}(i_1, ..., i_k)\, \boldsymbol{\Omega}(i_1, ..., i_k)$$

is a $k \times h$ matrix that constructs relation between $\boldsymbol{A}(i_1, ..., i_k)$ and the set of independent order statistics $X_{(i_1:h)}, ..., X_{(i_k:h)}$ in the larger set of size h. The distribution of $X_{[i]}$, conditional on the values of $i_1, ..., i_k$ is given by

$$F_{[i]}(x|i_1, ..., i_k) = \sum_{\iota=1}^{h} \boldsymbol{N}(i_1, ..., i_k)_{i\iota} F_{(\iota)}(x)$$

where $\boldsymbol{N}(i_1, ..., i_k)_{i\iota}$ is the (i, ι)-th entry of $\boldsymbol{N}(i_1, ..., i_k)$. When the contribution of all $\binom{h}{k}$ equally likely choices of the values of $i_1, ..., i_k$ the CDF of $X_{[i]}$ can then be written

$$F_{[i]}(x) = \sum_{\iota=1}^{h} p_{k,h}(i, \iota) F_{(\iota)}(x)$$

where $\boldsymbol{P}_{k,h} = (p_{k,h}(i,\iota))$ is the $k \times h$ matrix average

$$\boldsymbol{P}_{k,h} = \binom{h}{k}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le h} \boldsymbol{N}(i_1, ..., i_k)_{i\iota}$$

In our study, we assumed that $\boldsymbol{P}_{k,h}$ is a square matrix, so we use $\boldsymbol{P}$ and $p(i,\iota)$ instead of $\boldsymbol{P}_{k,h}$ and $p_{k,h}(i,\iota)$, respectively. For more details about Frey's one-parameter judgement ranking model, see Frey [7]. The matrix $\boldsymbol{P}$ can be estimated by using an R-function that is proposed by Ozturk [17] for any correlation coefficient (ρ), the set size (k) and the larger set size (h). For theoretical backgrounds of the R-function, see Ozturk [17]. In the following example, we illustrate the matrix $\boldsymbol{P}$.

Example 4.1. It is assumed that set size $k = 4$ and the units in the set are ranked perfectly. Then, the matrix $\boldsymbol{P}$ is as follows:

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

If the units in the set are ranked randomly, then the matrix $\boldsymbol{P}$ is as follows:

$$\boldsymbol{P} = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

Let us define that $X_{i[1]j}, X_{i[2]j}, ..., X_{i[k]j}$ be judgement order statistics in the i-th set, $i = 1, ..., k$ and $j = 1, ..., l$. Then, the EDF based on RSS [24] as follows.

$$\hat{F}^*_{RSS}(x) = \frac{1}{lk} \sum_{j=1}^{l} \sum_{i=1}^{k} I\big(X_{i[i]j} \le x\big) \tag{4.1}$$

On the other hand, the measured units in the steps of the PRSS are denoted by $X_{[r]}$, $X_{[s]}$ and $X_{[m]}$. Thus, the measured units in $PRSS_O$ and $PRSS_E$ are represented by

$$PRSS_O = \Big\{X_{1[r]j},\, X_{2[r]j},\, ...,\, X_{\frac{k-1}{2}[r]j},\, X_{m[m]j},\, X_{\frac{k+3}{2}[s]j},\, ...,\, X_{k-1[s]j},\, X_{k[s]j}\Big\}$$

and

$$PRSS_E = \Big\{X_{1[r]j},\, ...,\, X_{\frac{k}{2}[r]j},\, X_{\frac{k+2}{2}[s]j},\, ...,\, X_{k[s]j}\Big\},$$

respectively, where $m = (k+1)/2$ and $j = 1, ..., l$. The CDF estimators based on $PRSS_O$ and $PRSS_E$ are given by

$$\hat{F}^*_{PRSS_O}(x) = \frac{1}{lk}\left[\sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} I\big(X_{i[r]j} \le x\big) + \sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} I\Big(X_{\frac{k+1}{2}+i[s]j} \le x\Big) + \sum_{j=1}^{l} I\big(X_{m[m]j} \le x\big)\right] \tag{4.2}$$

and if set size is even,

$$\hat{F}^*_{PRSS_E}(x) = \frac{1}{lk}\left[\sum_{j=1}^{l}\sum_{i=1}^{\frac{k}{2}} I\big(X_{i[r]j} \le x\big) + \sum_{j=1}^{l}\sum_{i=1}^{\frac{k}{2}} I\Big(X_{\frac{k}{2}+i[s]j} \le x\Big)\right] \tag{4.3}$$

where $r \approx p(k+1)$, $s \approx q(k+1)$ and $m = \frac{k+1}{2}$ is the median ranked unit. The following proposition gives the properties of $\hat{F}^*_{PRSS_O}(x)$ and $\hat{F}^*_{PRSS_E}(x)$.

Proposition 4.1.

(**a**) *Using* $PRSS_O$

i. $E\Big(\hat{F}^*_{PRSS_O}(x)\Big) = \big(\frac{1}{2} - \frac{1}{2k}\big)\big(F_{[r]}(x) + F_{[s]}(x)\big) + \frac{1}{k}F_{[m]}(x)$,

ii. $V\Big(\hat{F}^*_{PRSS_O}(x)\Big) = \frac{1}{lk^2}\left[\begin{array}{c}\big(\frac{k-1}{2}\big)\big(F_{[r]}(x)\big(1-F_{[r]}(x)\big) + F_{[s]}(x)\big(1-F_{[s]}(x)\big)\big)+ \\ F_{[m]}(x)\big(1-F_{[m]}(x)\big)\end{array}\right]$;

(**b**) *Using* $PRSS_E$

i. $E\Big(\hat{F}^*_{PRSS_E}(x)\Big) = \frac{1}{2}\big(F_{[r]}(x) + F_{[s]}(x)\big)$,

ii. $V\Big(\hat{F}^*_{PRSS_E}(x)\Big) = \frac{1}{2lk}\big[F_{[r]}(x)\big(1-F_{[r]}(x)\big) + F_{[s]}(x)\big(1-F_{[s]}(x)\big)\big]$;

where

$$F_{[t]}(x) = \sum_{\iota=1}^{k} p(t,\iota)F_{(\iota)}(x), \qquad t = \{r,s,m\}.$$

The proof the Proposition 4.1 is the same as the proof of the Proposition 3.1. We gave an example in order to illustrate obtaining the distribution of judgement order statistics $F_{[t]}$. Also, we investigated the properties of $\hat{F}^*_{PRSS_O}(x)$ and $\hat{F}^*_{PRSS_E}(x)$ under random ranking case in this example. First, we give the following lemma that is noted by Dell and Clutter [6]. Detailed proof of this lemma was given by Presnell and Bohn [18].

Lemma 4.1. $\frac{1}{k}\sum_{i=1}^{k} F_{[i]}(x) = F(x), \ \forall x.$

Using this lemma, the results are provided in the following example.

Example 4.2. Let $\{X_{1[r]j}, X_{2[r]j}, ..., X_{\frac{k-1}{2}[r]j}, X_{m[m]j}, X_{\frac{k+3}{2}[s]j}, ..., X_{k-1[s]j}, X_{k[s]j}\}$ are obtained using $PRSS_O$ under random ranking case. Then, $p(t,\iota) = \frac{1}{k}$ in the matrix $\boldsymbol{P}$ for each $t = \{r,s,m\}$ and $\iota = 1, ..., k$. Thus, $F_{[t]}(x)$ is obtained according to Lemma 4.1:

$$F_{[t]}(x) = \sum_{\iota=1}^{k} \frac{1}{k}F_{(\iota)}(x) = F(x)$$

Straightforwardly, it can be seen that

$$E\Big(\hat{F}^*_{PRSS_O}(x)\Big) = F(x),$$

$$V\Big(\hat{F}^*_{PRSS_O}(x)\Big) = \frac{1}{n}F(x)(1-F(x)).$$

Besides, we have to note that the obtained results are not surprising. It means that $\hat{F}^*_{PRSS_O}(x)$ reduce to $\hat{F}(x)$ under random ranking case. Obviously, these results are the same for $\hat{F}^*_{PRSS_E}(x)$ as well.

Now, we investigated the performances of $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$ under the imperfect ranking. To construct imperfect ranking schemes, we take the correlation coefficients as $\boldsymbol{\rho} = \{0.90, 0.75, 0.50\}$. The matrix $\boldsymbol{P}_\upsilon$, $\upsilon = 1, 2, 3$ corresponding to each correlation coefficient are estimated using Ozturk's R-function. When $k = 3$, the estimated matrices are

$$\boldsymbol{P}_1 = \begin{bmatrix} 0.841 & 0.151 & 0.008 \\ 0.151 & 0.698 & 0.151 \\ 0.008 & 0.151 & 0.841 \end{bmatrix},$$

$$\boldsymbol{P}_2 = \begin{bmatrix} 0.762 & 0.210 & 0.028 \\ 0.210 & 0.580 & 0.210 \\ 0.028 & 0.210 & 0.762 \end{bmatrix},$$

$$\textit{and} \quad \boldsymbol{P}_3 = \begin{bmatrix} 0.555 & 0.303 & 0.142 \\ 0.303 & 0.395 & 0.303 \\ 0.142 & 0.303 & 0.555 \end{bmatrix},$$

for $\rho = 0.90$, $\rho = 0.75$ and $\rho = 0.50$, respectively. These matrices are estimated for $k = 4$, $k = 5$ and $k = 6$ as well. Bias for $\hat{F}^*_{PRSS_O}(x)$ and $\hat{F}^*_{PRSS_E}(x)$ are obtained by using Equations (4.4) and (4.5). Figure 3 gives bias for the CDF estimators based on PRSS with $p = 0.1$ and $p = 0.4$, respectively. For any ρ, these EDF estimators are unbiased as $F(x)$ gets closer to 1, 0.5 and 0. Also, the bias increases as k increases except for $F(x) = 0.5$. It can be seen that the biases decrease as ρ decreases. This is a result of the Example 4.2:

$$Bias[\hat{F}^*_{PRSS_O}(x)] = F(x) - E(\hat{F}^*_{PRSS_O}(x)), \tag{4.4}$$

$$Bias[\hat{F}^*_{PRSS_E}(x)] = F(x) - E(\hat{F}^*_{PRSS_E}(x)). \tag{4.5}$$

Besides, relative efficiencies (RE) of $\hat{F}_{PRSS_O}(x)$ and $\hat{F}_{PRSS_E}(x)$ with respect to $\hat{F}_{SRS}(x)$ are described as

$$RE[\hat{F}^*_{PRSS_O}(x), \hat{F}_{SRS}(x)] = \frac{V(\hat{F}_{SRS}(x))}{MSE(\hat{F}^*_{PRSS_O}(x))},$$

$$RE[\hat{F}^*_{PRSS_E}(x), \hat{F}_{SRS}(x)] = \frac{V(\hat{F}_{SRS}(x))}{MSE(\hat{F}^*_{PRSS_E}(x))}.$$

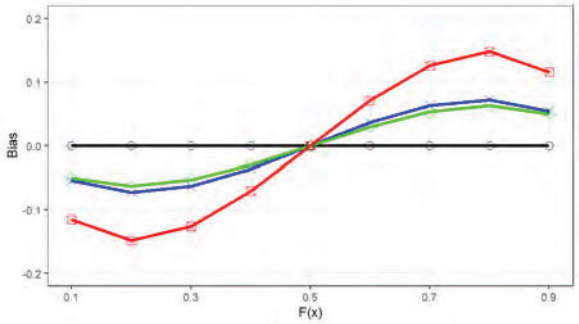
REs are given by Figure 4 for $p = 0.1$ and $p = 0.4$, respectively. For any ρ, it is seen that the REs peak on the middle of the distribution function when $p = 0.1$. Also, the REs increase while the set size increases. On the other hand, the REs are higher on the tails of the distribution function when $p = 0.4$. Also, the REs are almost equal to or larger than 1 for any $F(x)$ and ρ when $k = 3, 4, 5, 6$ and $p = 0.4$.

Table 2 gives REs of EDFs using PRSS when $F(x) = 0.1$ and $F(x) = 0.5$ relative to RSS. The REs are obtained by using the following equations.

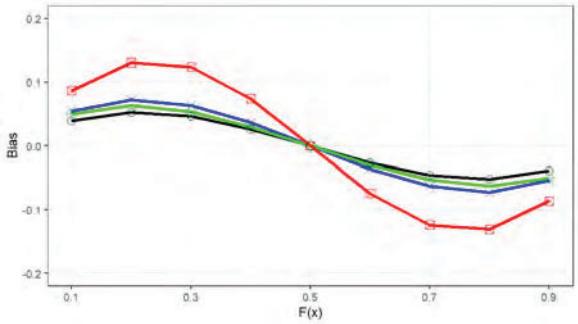
$$RE[\hat{F}^*_{PRSS_O}(x), \hat{F}^*_{RSS}(x)] = \frac{V(\hat{F}^*_{RSS}(x))}{MSE(\hat{F}^*_{PRSS_O}(x))},$$

and

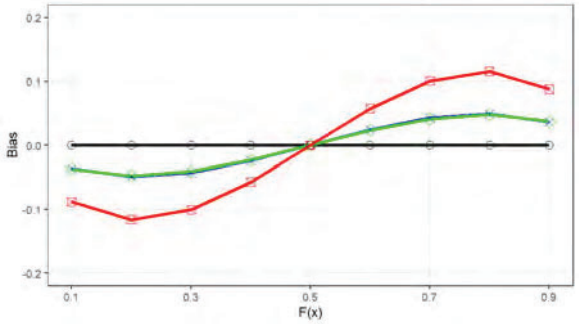
$$RE[\hat{F}^*_{PRSS_E}(x), \hat{F}^*_{RSS}(x)] = \frac{V(\hat{F}^*_{RSS}(x))}{MSE(\hat{F}^*_{PRSS_E}(x))}.$$



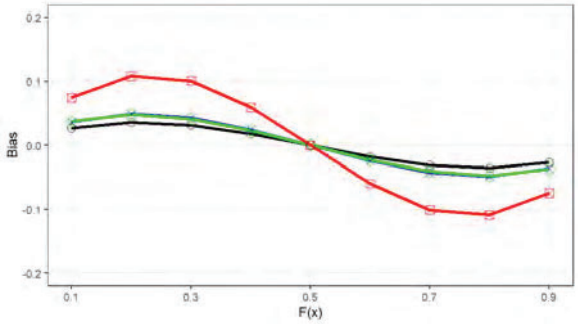
(a) When $p = 0.1$, the bias of EDFs for $\rho = 0.90$.



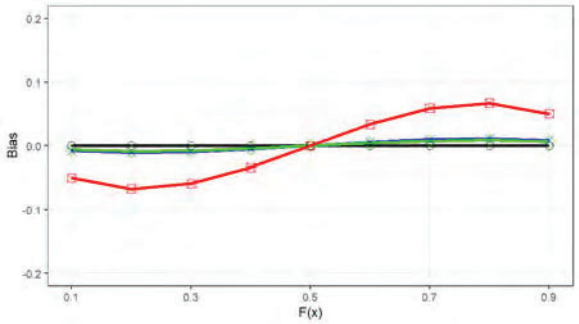
(b) When $p = 0.4$, the bias of EDFs for $\rho = 0.90$.



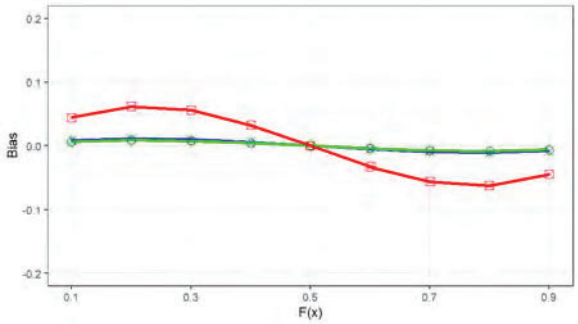
(c) When $p = 0.1$, the bias of EDFs for $\rho = 0.75$.



(d) When $p = 0.4$, the bias of EDFs for $\rho = 0.75$.

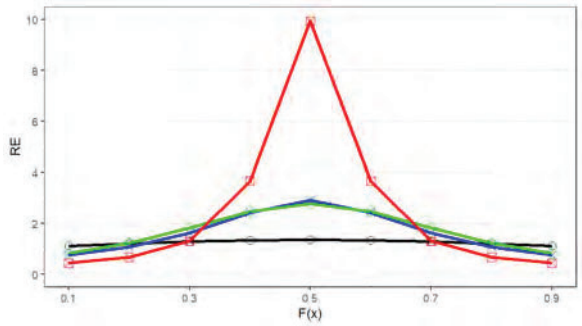


(e) When $p = 0.1$, the bias of EDFs for $\rho = 0.50$.

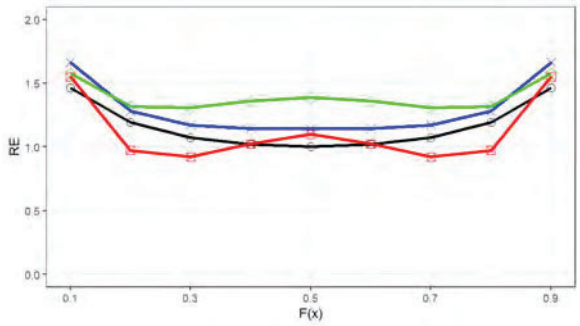


(f) When $p = 0.4$, the bias of EDFs for $\rho = 0.50$.

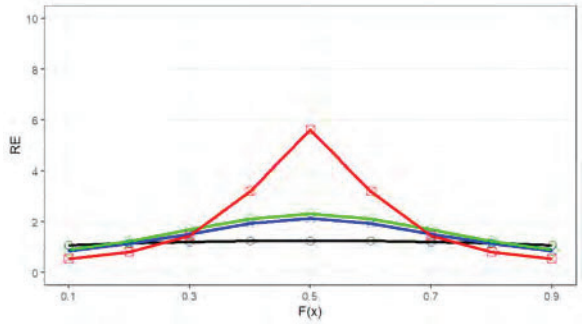
Figure 3: Bias for $\hat{F}_{PRSS_O}$ and $\hat{F}_{PRSS_E}$ where black, blue, green and red curves are $k = 3$, $k = 4$, $k = 5$ and $k = 6$, respectively.



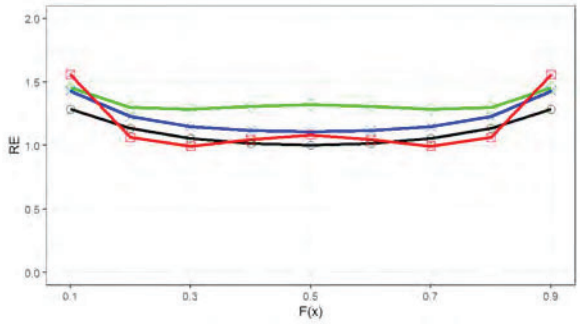
(a) When $p = 0.1$, the REs of EDFs for $\rho = 0.90$.



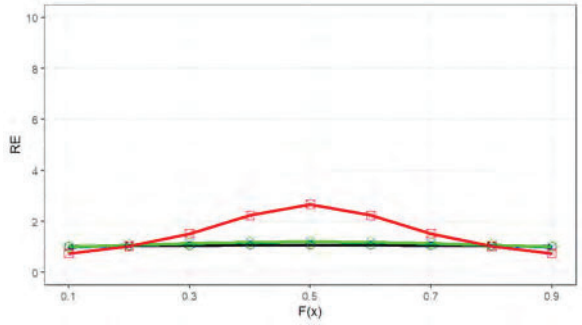
(b) When $p = 0.4$, the REs of EDFs for $\rho = 0.90$.



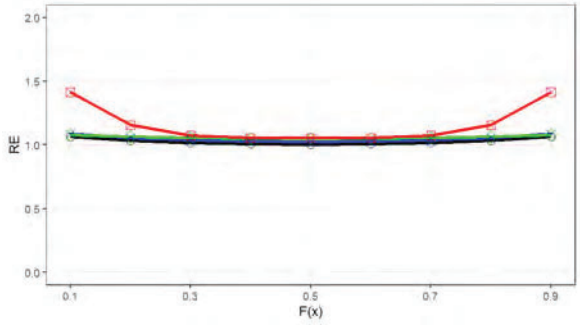
(c) When $p = 0.1$, the REs of EDFs for $\rho = 0.75$.



(d) When $p = 0.4$, the REs of EDFs for $\rho = 0.75$.



(e) When $p = 0.1$, the REs of EDFs for $\rho = 0.50$.



(f) When $p = 0.4$, the REs of EDFs for $\rho = 0.50$.

Figure 4: REs for $\hat{F}_{PRSS_O}$ and $\hat{F}_{PRSS_E}$ where black, blue, green and red curves are $k = 3$, $k = 4$, $k = 5$ and $k = 6$, respectively.

Table 2 shows that even if $\rho = 0.5$, the gain in efficiency from EDFs using PRSS with $p = 0.4$ (and with $p = 0.1$) are substantial when $F(x) = 0.1$ (and when $F(x) = 0.5$).

Table 2: The REs of the EDF estimators based on PRSS with respect to RSS.

		$F(x)=0.1$		$F(x)=0.5$	
ρ	k	$p=0.1$	$p=0.4$	$p=0.1$	$p=0.4$
0.9	3	1.000	1.312	1.000	0.740
	4	0.629	1.379	1.782	0.695
	5	0.647	1.233	1.504	0.749
	6	0.323	1.128	4.784	0.527
0.75	3	1.000	1.185	1.000	0.798
	4	0.749	1.251	1.461	0.760
	5	0.738	1.204	1.379	0.785
	6	0.434	1.231	2.993	0.578
0.5	3	1.000	1.036	1.000	0.936
	4	0.952	1.046	1.091	0.923
	5	0.962	1.036	1.070	0.939
	6	0.651	1.231	1.775	0.696

The Proposition 4.2 is needed to study some asymptotic inference about the expected value of the estimators, $\hat{F}^*_{PRSS_O}(x)$ and $\hat{F}^*_{PRSS_E}(x)$.

Proposition 4.2. *For fixed k and $l \to \infty$, the following results are obtained:*

(**a**) $\dfrac{\hat{F}^*_{PRSS_O}(x)-E\left(\hat{F}^*_{PRSS_O}(x)\right)}{\sqrt{V\left(\hat{F}^*_{PRSS_O}(x)\right)}}$ *converges in distribution to* $N(0,1)$;

(**b**) $\dfrac{\hat{F}^*_{PRSS_E}(x)-E\left(\hat{F}^*_{PRSS_E}(x)\right)}{\sqrt{V\left(\hat{F}^*_{PRSS_E}(x)\right)}}$ *converges in distribution to* $N(0,1)$.

The proof of the Proposition 4.2 is similar to proof of Proposition 3.2.

5. INFERENCES ABOUT $F(x)$

In this section, we now consider a pointwise estimate of $F(x)$. It supposed that we interest with the proportion, $F(x)$ of population below a specified value X. We know that $100(1-\alpha)\%$ confidence interval for $F(x)$ using SRS is as follows:

$$\hat{F}_{SRS}(x) \pm Z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{F}_{SRS}(x))}$$

where $Z_{\frac{\alpha}{2}}$ is the upper quantile of the standard normal distribution and

$$\hat{V}(\hat{F}_{SRS}(x)) = \frac{1}{n-1}\hat{F}_{SRS}(x)\Big(1-\hat{F}_{SRS}(x)\Big).$$

Also, Stokes and Sager [24] gave a $100(1-\alpha)\%$ for $F(x)$ using RSS:

$$\hat{F}_{RSS}(x) \pm Z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{F}_{RSS}(x))}$$

where

$$\hat{V}(\hat{F}_{RSS}(x)) = \frac{1}{(l-1)k}\sum_{i=1}^{k}\hat{F}_{(i)}(x)\Big(1-\hat{F}_{(i)}(x)\Big).$$

According to Proposition 3.2, an approximate $100(1-\alpha)\%$ confidence intervals can be constructed when l is larger. For $\hat{F}_{PRSS_O}(x)$, confidence interval of $F(x)$ can be obtained as

$$p\left(Z_{\frac{\alpha}{2}} \leq \frac{\hat{F}_{PRSS_O}(x) - E(\hat{F}_{PRSS_O}(x))}{\sqrt{\hat{V}(\hat{F}_{PRSS_O}(x))}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha, \tag{5.1}$$

where

$$\begin{aligned}\hat{V}(\hat{F}_{PRSS_O}(x)) = \frac{1}{(l-1)k^2}\Bigg[&\left(\frac{k-1}{2}\right)\hat{F}_{(r)}(x)(1-\hat{F}_{(r)}(x)) \\ &+ \left(\frac{k-1}{2}\right)\hat{F}_{(s)}(x)(1-\hat{F}_{(s)}(x)) + \hat{F}_{(m)}(x)(1-\hat{F}_{(m)}(x))\Bigg].\end{aligned}$$

By solving the Equation (5.1) for $E(\hat{F}_{PRSS_O}(x))$, the limits are obtained.

$$Lower\ Bound(LB) = \hat{F}_{PRSS_O}(x) - Z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{F}_{PRSS_O}(x))},$$

and

$$Upper\ Bound(UB) = \hat{F}_{PRSS_O}(x) + Z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{F}_{PRSS_O}(x))}.$$

Thus, $100(1-\alpha)\%$ confidence interval of $F(x)$ can be found by solving the following equations, numerically or any suitable method such as Newton Raphson.

$$\begin{aligned}2LB &= \frac{1}{k}(k-1)\big(F_{(r)}(x)+F_{(s)}(x)\big) + 2F_{(m)}(x) \\ &= \Psi(F),\end{aligned} \tag{5.2}$$

and

$$\begin{aligned}2UL &= \frac{1}{k}(k-1)\big(F_{(r)}(x)+F_{(s)}(x)\big) + 2F_{(m)}(x) \\ &= \Psi(F).\end{aligned} \tag{5.3}$$

For confidence interval of $F(x)$ based on $\hat{F}_{PRSS_E}(x)$,

$$p\left(Z_{\frac{\alpha}{2}} \leq \frac{\hat{F}_{PRSS_E}(x) - E(\hat{F}_{PRSS_E}(x))}{\sqrt{\hat{V}(\hat{F}_{PRSS_E}(x))}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha, \tag{5.4}$$

where

$$\hat{V}(\hat{F}_{PRSS_E}(x)) = \frac{1}{2(l-1)k}\Big[\hat{F}_{(r)}(x)(1-\hat{F}_{(r)}(x)) + \hat{F}_{(s)}(x)(1-\hat{F}_{(s)}(x))\Big].$$

Thus, the limits are obtained as

$$LB = \hat{F}_{PRSS_E}(x) - Z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{F}_{PRSS_E}(x))},$$

and

$$UB = \hat{F}_{PRSS_E}(x) + Z_{\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{F}_{PRSS_E}(x))}.$$

$100(1-\alpha)\%$ confidence interval of $F(x)$ can be found by solving the following equations:

$$2LB = F_{(r)}(x) + F_{(s)}(x) = \Psi(F), \tag{5.5}$$

and

$$2UL = F_{(r)}(x) + F_{(s)}(x) = \Psi(F). \tag{5.6}$$

Note that $\Psi(F)$ is increasing function in $F(x)$ so the solutions of the Equations (5.2), (5.3), (5.5) and (5.6) should be unique. Similarly, confidence intervals are obtained using $\hat{F}^*_{PRSS_O}(x)$ and $\hat{F}^*_{PRSS_E}(x)$.

5.1. A real data application

In the literature, the distribution function estimators are applied to real data such as bilirubin level [19], lung cancer [28] and airquality [27]. The number of case studies can be increased. In the case studies, it can be seen that some quantiles are important hence the probabilities corresponding to them are substantial as well. Thus, if we can estimate the distribution function, these probabilities can also be estimated.

In this section, we consider body mass index data (BMI) to give an illustrative example. BMI is a measure for indicating nutritional status in adults. BMI is frequently used to screen for weight categories that may lead to health problems. A table that includes the weight categories was reported by World Health Organization (WHO), `http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi` and this categories are given by Table 3.

Table 3: The weight categories.

BMI	Nutritional status
Below 18.5	Underweight
18.5−24.9	Normal weight
25.0−29.9	Pre-obesity
30.0−34.9	Obesity class I
35.0−39.9	Obesity class II
Above 40	Obesity class III

According to WHO, the health problems caused by obesity are as follows: premature death, cardiovascular diseases, high blood pressure, osteoarthritis, some cancers and diabetes.

Orginal data includes 500 adult people (255 of 500 are women) and four variables such as gender, height (m), weight (kg) and index (0: extremely weak, 1: weak, 2: normal, 3: overweight, 4: obesity and 5: extreme obesity). This data can be available in `https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex`. However, we assume a population that includes 255 women and their measurements such

as height (m) and weight (kg) in our study. Note that we limited the population size as 255 to give sample observations. Thus, we aimed to illustrate the application, clearly. Also, it is supposed that the proportion of women in the Obesity class III is close to 0.5, $1-F(40)\approx 0.5$. Therefore, using PRSS with $p=0.1$ is appropriate in this case. From this population, $n=100$ observations are selected using PRSS with $p=0.1$. To obtain PRSS, we take the set size and the number of cycles as $k=5$ and $l=20$, respectively.

In the process PRSS, 25 observations are first selected at random among 255 women in j-th cycle, $j=1,...,20$. Then, the 25 observations are assigned into 5 sets at random. Ranking the BMI of the 25 observations may be performed by subjective ranking or according to a concomitant variable such as height of the observations. Also, it is assumed that ranking is almost perfect. The ranked sets are given as follows.

Table 4: Selected units in PRSS for j-th cycle, $j=1,...,20$.

Set	Ranked Units	Measured Units
S_1	$\boldsymbol{X_{1[1]j}} \le X_{1[2]j} \le X_{1[3]j} \le X_{1[4]j} \le X_{1[5]j}$	$X_{1[1]j}$
S_2	$\boldsymbol{X_{2[1]j}} \le X_{2[2]j} \le X_{2[3]j} \le X_{2[4]j} \le X_{2[5]j}$	$X_{2[1]j}$
S_3	$X_{3[1]j} \le X_{3[2]j} \le \boldsymbol{X_{3[3]j}} \le X_{3[4]j} \le X_{3[5]j}$	$X_{3[3]j}$
S_4	$X_{4[1]j} \le X_{4[2]j} \le X_{4[3]j} \le X_{4[4]j} \le \boldsymbol{X_{4[5]j}}$	$X_{4[5]j}$
S_5	$X_{5[1]j} \le X_{5[2]j} \le X_{5[3]j} \le X_{5[4]j} \le \boldsymbol{X_{5[5]j}}$	$X_{5[5]j}$

In the sets, bold faced units represent the measured BMIs of 5 observations among 25 observations. For the first cycle, the measured BMIs are $X_{1[1]1}=18.52$, $X_{2[1]1}=12.75$, $X_{3[3]1}=32.45$, $X_{4[5]1}=52.89$ and $X_{4[5]1}=66.66$. These BMIs are given in the first row of Table 5. $1-\hat{F}_{PRSS_O}(40)=0.41$ is obtained according to the sample. Also, 95% confidence interval of $1-F(40)\approx 0.5$ is $(0.35, 0.46)$.

6. CONCLUSION

In this study, PRSS procedure is considered to estimate the distribution function. Properties of the EDF using PRSS are investigated. We examined how well the estimator performs in comparison with its SRS and RSS counterparts. Finally, we can summarize the following remarks:

1. Whether the ranking is perfect or not, the EDFs based on PRSS are unbiased as $F(x)$ gets closer to 1, 0.5 and 0.

2. Compared with $\hat{F}_{SRS}(x)$, the EDFs based on PRSS are more efficient under perfect and imperfect ranking.

3. If there is a known prior information that the value of $F(x)$ gets closer to 0.1, PRSS with $p=0.4$ can be preferred instead of RSS whether the ranking is perfect or not.

4. Also, PRSS with $p=0.1$ can be preferred instead of RSS when $F(x)$ is close to 0.5.

5. As in our application for BMI data, PRSS with $p = 0.1$ is recommended when estimating for the center of the distribution.
6. Also, it is suggested to use PRSS with $p = 0.4$ when estimating the extremes of the distribution.
7. In many studies on EDF estimators based on RSS and its modifications, theoretical results are presented for perfect ranking case while empirical results are presented for imperfect ranking case. Empirical results are obtained by running Monte Carlo simulations in the studies. Unlike the other studies, the present paper shows that the proposed EDF estimator can be examined theoretically by using Frey [7]'s ranking error model even in the case of imperfect ranking.

As a future work, the moment-based (MB) and maximum likelihood (ML) estimators of the CDF can be considered. A comparable study of the MB, ML and the EDF estimators based on PRSS can be meaningful. The authors continue to work towards this goal.

A. APPENDIX

Proof of Theorem 2.1

To prove this theorem, we follow the Proof of Lemma 2.1 in Samawi and Al-Sagheer [19] and the Proof of Theorem 1 in Stokes and Sager[24].

(**1**) Units in $PRSS_O$ are sampled from specific groups. It is assumed that $t_r = \frac{(k-1)l}{2}$ observations comes from $f_{(r)}(x)$, $t_s = \frac{(k-1)l}{2}$ from $f_{(s)}(x)$ and $t_m = l$ from $f_{(m)}(x)$, where $f_{(m)}(x)$ is density function of m-th order statistic. Note that $t_1 = \cdots = t_{r-1} = 0$, $t_{r+1} = \cdots = t_{m-1} = 0$, $t_{m+1} = \cdots = t_{s-1} = 0$ and $t_{s+1} = \cdots = t_k = 0$. This is accomplished by first randomly select R from $1, ..., k$ with replacement and if $r = 1$, $r = k$ or $r = m$ then observe Y according to $F_r(x)$, otherwise reject r. In SRS the order in which the groups are sampled is random, by rearranging and relabeling, a realization $(y_1, ..., y_{kl})$ of $(Y_1, ..., Y_{kl})$ becomes $(Z_{r1}, ..., Z_{r\frac{(k-1)l}{2}}, Z_{m1}, ..., Z_{ml}, Z_{s1}, ..., Z_{s\frac{(k-1)l}{2}})$ the groups $\{Z_{ij}, Z_{mj'}; i = r, s; j = 1, ..., \frac{(k-1)l}{2}; j' = 1, ..., l\}$. It is necessary to specify a consistent order for the units of the $PRSS_O$ and SRS to compare their distributions logically. Otherwise, because of the arbitrariness of listing order, a coordinate wise of PDF's or CDF's between $PRSS_O$ and SRS might imply unequal distributions, although the only difference would be a permutation of coordinates. Given

$$\underline{T} = \left(0, ..., 0, t_r = \frac{(k-1)l}{2}, 0, ..., 0, t_m = l, 0..., 0, \ t_s = \frac{(k-1)l}{2}, 0, ..., 0\right)$$

and $P(\underline{T}=t_i)=\frac{1}{k}, i=1, ..., k$ then, there are $\frac{(kl)!}{t_r!\cdots t_m!\cdots t_s!} = \frac{(kl)!}{\left(\left(\frac{(k-1)l}{2}\right)!\right)^2 l!}$ rearrangements of Y yielding the same Z. So the conditional CDF of Y given $\underline{T} = \underline{t}$ is

$$\frac{1}{P(\underline{T} = \underline{t})} P\Big\{ Z_{r1} \le a_{r1}, ..., Z_{r\frac{(k-1)l}{2}} \le a_{r\frac{(k-1)l}{2}}, Z_{m1} \le a_{m1}, ..., Z_{ml} \le a_{ml}, Z_{s1} \le a_{s1}, ..., Z_{s\frac{(k-1)l}{2}} \le a_{s\frac{(k-1)l}{2}}; \underline{T}\Big\} =$$

$$= \frac{1}{\frac{(kl)!}{(t_r!\cdots t_m!\cdots t_s!)}\left(\frac{1}{k}\right)^{t_r}\cdots\left(\frac{1}{k}\right)^{t_m}\cdots\left(\frac{1}{k}\right)^{t_s}}$$

$$\times \sum \left[\prod_{i=1}^{\frac{(k-1)l}{2}} \left(F_{(r)}(a_{ri}) \times \frac{1}{k}\right)\left(F_{(s)}(a_{si}) \times \frac{1}{k}\right) \times \prod_{i'=1}^{l}\left(F_{(m)}(a_{mi'}) \times \frac{1}{k}\right)\right]$$

where the sum is over all rearrangements of Y consistent with $\underline{T} = \underline{t}$. So

$$\sum \frac{\prod_{i=1}^{\frac{(k-1)l}{2}} \left(F_{(r)}(a_{ri}) \times F_{(s)}(a_{si})\right) \prod_{i'=1}^{l}\left(F_{(m)}(a_{mi'})\right)}{\frac{(kl)!}{\left(\left(\frac{(k-1)l}{2}\right)!\right)^2 l!}}$$

$$= \prod_{i=1}^{\frac{(k-1)l}{2}} \left(F_{(r)}(a_{ri}) \times F_{(s)}(a_{si})\right) \prod_{i'=1}^{l}\left(F_{(m)}(a_{mi'})\right).$$

(**2**) It is assumed that $t_r = \frac{lk}{2}$ observations come from $f_{(r)}(x)$ and $t_s = \frac{lk}{2}$ from $f_{(s)}(x)$, where $f_{(r)}(x)$ and $f_{(s)}(x)$ are density functions of r-th and s-th order statistics, respectively. This proof follows from the part (1).

Proof of Proposition 3.1

(**a**) For $\hat{F}_{PRSS_O}(x)$,

i.

$$E\Big(\hat{F}_{PRSS_O}(x)\Big) = \frac{1}{lk}\Bigg[\sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} E\big(I\big(X_{i(r)j} \le x\big)\big) + \sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} E\Big(I\Big(X_{\frac{k+1}{2}+i(s)j} \le x\Big)\Big) + \sum_{j=1}^{l} E\big(I\big(X_{m(m)j} \le x\big)\big)\Bigg]$$

$I\big(X_{i(r)j} \le x\big)$, $I\Big(X_{\frac{k+1}{2}+i(s)j} \le x\Big)$ and $I\big(X_{m(m)j} \le x\big)$ have Bernoulli distributions with parameters $F_{(r)}(x)$, $F_{(s)}(x)$ and $F_{(m)}(x)$, respectively. Therefore,

$E\big(I\big(X_{i(r)j} \le x\big)\big) = F_{(r)}(x)$,

$E\Big(I\Big(X_{\frac{k+1}{2}+i(s)j} \le x\Big)\Big) = F_{(s)}(x)$ and

$E\big(I\big(X_{m(m)j} \le x\big)\big) = F_{(m)}(x)$.

Thus,

$$E\Big(\hat{F}_{PRSS_O}(x)\Big) = \left(\frac{1}{2} - \frac{1}{2k}\right)\big(F_{(r)}(x) + F_{(s)}(x)\big) + \frac{1}{k}F_{(m)}(x).$$

ii.

$$V\Big(\hat{F}_{PRSS_O}(x)\Big) = \frac{1}{lk}\Bigg[\sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} V\big(I\big(X_{i(r)j} \le x\big)\big) + \sum_{j=1}^{l}\sum_{i=1}^{\frac{k-1}{2}} V\Big(I\Big(X_{\frac{k+1}{2}+i(s)j} \le x\Big)\Big) + \sum_{j=1}^{l} V\big(I\big(X_{m(m)j} \le x\big)\big)\Bigg]$$

Since $I\big(X_{i(r)j} \le x\big)$, $I\Big(X_{\frac{k+1}{2}+i(s)j} \le x\Big)$ and $I\big(X_{m(m)j} \le x\big)$ have Bernoulli distribution, variance of these indicator functions are given bellow:

$V\big(I\big(X_{i(r)j} \le x\big)\big) = F_{(r)}(x)\big(1 - F_{(r)}(x)\big)$,

$$V(I(X_{\frac{k+1}{2}+i(s)j} \leq x)) = F_{(s)}(x)(1 - F_{(s)}(x)),$$

$$V\big(I\big(X_{m(m)j} \leq x\big)\big) = F_{(m)}(x)\big(1 - F_{(m)}(x)\big).$$

Thus, variance of the estimator can be obtained:

$$\begin{aligned} V\Big(\hat{F}_{PRSS_O}(x)\Big) \;=\; & \frac{1}{lk^2}\Bigg[\left(\frac{k-1}{2}\right)F_{(r)}(x)\big(1 - F_{(r)}(x)\big) \\ & + \left(\frac{k-1}{2}\right)F_{(s)}(x)\big(1 - F_{(s)}(x)\big) \\ & + F_{(m)}(x)\big(1 - F_{(m)}(x)\big)\Bigg]. \end{aligned}$$

(**b**) $E\Big(\hat{F}_{PRSS_E}(x)\Big)$ and $V\Big(\hat{F}_{PRSS_E}(x)\Big)$ can be proved by using the same steps in Proof (a).

Proof of Proposition 3.2

Following Samawi and Al-Sagheer[19] and Kim *et al.* [12],

(**a**) Let $Z_j = \frac{1}{k}\left[\sum_{i=1}^{\frac{k-1}{2}}\Big(I(X_{i(r)j} \leq x) + I\Big(X_{\frac{k+1}{2}+i(s)j} \leq x\Big)\Big) + I(X_{m(m)j} \leq x)\right],\ j = 1, ..., l.$ Since Z_j are independent and identically with finite mean and variance, then based on Central Limit Theorem

$$\left(\frac{\bar{Z} - E(Z_j)}{\left(\frac{var(Z_j)}{l}\right)^{1/2}}\right) \xrightarrow{D} N(0,1)$$

(**b**) Similarly, this part can be proved by assuming

$$Z_j = \frac{1}{k}\sum_{i=1}^{\frac{k}{2}}\Big(I(X_{i(r)j} \leq x) + I\Big(X_{\frac{k}{2}+i(s)j} \leq x\Big)\Big).$$

Percentile ranked set sample

Table 5: Sample observations that are obtained using PRSS.

	PRSS				
l	1^{st}	1^{st}	5^{th}	5^{th}	3^{rd}
1	18.52	12.75	52.89	66.66	32.45
2	23.59	28.20	43.17	53.01	37.57
3	12.75	20.90	66.66	40.75	39.21
4	21.37	20.96	43.11	68.96	30.48
5	16.38	28.07	57.96	57.70	32.42
6	29.17	30.64	52.89	66.66	28.67
7	20.90	22.65	67.06	52.89	35.58
8	25.98	17.43	43.56	57.70	32.42
9	22.63	33.96	44.63	71.93	32.15
10	24.12	22.45	57.96	54.86	30.42
11	20.02	28.07	48.15	59.49	33.77
12	17.43	27.35	68.41	59.69	35.58
13	18.34	16.04	51.17	55.66	32.42
14	24.12	25.46	44.90	53.01	44.79
15	28.07	20.52	32.69	59.94	39.44
16	12.75	35.29	67.94	78.85	49.34
17	33.88	17.09	59.84	71.93	43.56
18	21.37	20.52	39.06	78.85	52.80
19	22.10	32.15	43.17	52.26	39.68
20	23.23	26.40	36.95	63.38	23.59

ACKNOWLEDGMENTS

The authors thank to Professor Omer Ozturk for providing the R function that is used in computation of ranking error probabilities. Also, the authors are grateful to the referees and the editor for helpful comments.

REFERENCES

[1] Abu-Dayyeh, W.A.; Samawi, H.M. and Bani-Hani L.A. (2002). On distribution function estimation using double ranked set samples with application, *Journal of Modern Applied Statistical Methods*, **1**(2), 443–451.

[2] Al-Omari, A.I. (2016). Quartile ranked set sampling for estimating the distribution function, *Journal of the Egyptian Mathematical Society*, **24**(2), 303–308.

[3] Al-Omari, A.I. and Bouza, C.N. (2014). Review of ranked set sampling: modifications and applications, *Revista Investigación Operacional*, **35**(3), 215–240.

[4] Al-Subh, S.A.; Alodat, M.T.; Ibrahim, K. and Jemain, A.A. (2009). EDF goodness of fit tests of logistic distribution under selective order statistics, *Pakistan Journal of Statistics*, **25**(3), 265–274.

[5] Bohn, L.L. and Wolfe, D.A. (1994). The effect of imperfect judgment rankings on properties of procedures based on the ranked-set samples analog of the Mann-Whitney-Wilcoxon statistic, *Journal of the American Statistical Association*, **89**(425), 168–176.

[6] Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background, *Biometrics*, **28**(2), 545–555.

[7] Frey, J. (2007a). New imperfect ranking models for ranked set sampling, *Journal of Statistical Planning and Inference*, **137**(4), 1433–1445.

[8] Frey, J. (2007b). Distribution-free statistical intervals via ranked set sampling, *The Canadian Journal of Statistics*, **35**(4), 585–596.

[9] Jemain, A.A. and Al-Omari, A.I. (2006). Double percentile ranked set samples for estimating the population mean, *Advances and Applications in Statistics*, **6**(3), 261–276.

[10] Jemain, A.A. and Al-Omari, A.I. (2007). Multistage percentile ranked set samples, *Advances and Applications in Statistics*, **7**(1), 127–139.

[11] Kaur, A.; Patil, G.P.; Sinha, A.K. and Taillie, C. (1995). Ranked set sampling: an annotated bibliography, *Environmental and Ecological Statistics*, **2**, 25–54.

[12] Kim, D.H.; Kim, D.W. and Kim, G.H. (2005). On the estimation of the distribution function using extreme median ranked set sampling, *Journal of the Korean Data Analysis Society*, **7**, 429–439.

[13] McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets, *Australian Journal of Agricultural Research*, **3**(4), 385–390.

[14] Muttlak, H.A. (2003). Modified ranked set sampling methods, *Pakistan Journal of Statistics*, **19**, 315–323.

[15] Neerchal, N.K.; Sinha, B.K. and Lacayo, H. (1998). Ranked set sampling from a dichotomous population, *Journal of Applied Statistical Science*, **11**(1), 83–90.

[16] Nazari, S.; Jafari Jozani, M. and Kharrati-Kopaei, M. (2016). On distribution function estimation with partially rank-ordered set samples: estimating mercury level in fish using length frequency data, *Statistics*, **50**(6), 1387–1410.

[17] Ozturk, O. (2008). Inference in the presence of ranking error in ranked set sampling, *The Canadian Journal of Statistics*, **36**(4), 557–594.

[18] Presnell, B. and Bohn, L.L. (1999). U-Statistics and imperfect ranking in ranked set sampling, *Nonparametric Statistics*, **10**(2), 111–126.

[19] Samawi, H.M. and Al-Sagheer, A.M. (2001). On the estimation of the distribution function using extreme and median ranked set sampling, *Biometrical Journal*, **43**(3), 357–373.

[20] Sevil, Y.C. and Yildiz, T.O. (2017). Power comparison of the Kolmogorov–Smirnov test under ranked set sampling and simple random sampling, *Journal of Statistical Computation and Simulation*, **87**(11), 2175–2185.

[21] Sevil, Y.C. and Yildiz, T.O. (2022). Design-based estimators of the distribution function in ranked set sampling with an application, *Statistics*, **56**(4), 891–918.

[22] Stokes, S.L. (1980a). Inferences on the correlation coefficient in bivariate normal populations from ranked set sampling, *Journal of the American Statistical Association*, **75**(372), 989–995.

[23] Stokes, S.L. (1980b). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, **36**(1), 35–42.

[24] Stokes, S.L. and Sager, T.W. (1988). Characterization of a ranked-set sample with application to estimating distribution function, *Journal of the American Statistical Association*, **83**(402), 374–381.

[25] Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on sample stratified by means of ordering, *Annals of the Institute of Statistical Mathematics*, **20**, 1–31.

[26] Yildiz, T.O. and Sevil, Y.C. (2018). Performances of some goodness-of-fit tests for sampling designs in ranked set sampling, *Journal of Statistical Computation and Simulation*, **88**(9), 1702–1716.

[27] Yildiz, T.O. and Sevil, Y.C. (2019). Empirical distribution function estimators based on sampling designs in a finite population using single auxiliary variable, *Journal of Applied Statistics*, **46**(16), 2962–2974.

[28] Zamanzade, E. (2019). EDF-based tests of exponentiality in pair ranked set sampling, *Statistical Papers*, **60**(6), 2141–2159.

[29] Zamanzade, E. and Mahdizadeh, M. (2020). Using ranked set sampling with extreme ranks in estimating the the population proportion, *Statistical Methods in Medical Research*, **29**(1), 165–177.

[30] Zamanzade, E. and Wang, X. (2018). Proportion estimation in ranked set sampling in the presence of tie information, *Computational Statistics*, **33**(3), 1349–1366.

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 63–76
https://doi.org/10.57805/revstat.v21i1.393

Modelling Pages Left Blank in University Examination: A Resolution in Higher Education Process

Authors: Suman K. Ghosh
– Alliance School of Applied Mathematics, Alliance University, India
suman.ghosh.2006@gmail.com

Subhradev Sen ✉
– Alliance School of Applied Mathematics, Alliance University, India
subhradev.stat@gmail.com

Received: April 2020 Revised: July 2021 Accepted: July 2021

Abstract:

- Trees are the main sources of paper production, in most of the cases, as far as the intellectual usages are concerned. However, our planet is lacking in that particular natural resource due to rapid growth of population, urbanization, and increased pollution, more importantly non-judicial utilization of such kind. Indian education sectors (schools, colleges, universities) utilize a major part in consumption of papers as a classical practice for conducting examinations and other documentation activities. Our attempt in this article is to investigate and provide an optimal estimate of the number of pages actually required in answer booklet in higher education sector. Truncated Poisson distribution is found to be the best fit for the data on number of pages left blank in an answer booklet after conduction of semester end examinations. To predict the outcome based on various factors such as, lines per pages, words per line, types of examinations etc. suitable regression modelling is performed. A real data set, collected over a period of one month, is been analysed to illustrate the methods and conclusion is accomplished in the direction of cost reduction, saving of papers, and in turn, logical uses of natural resource to protect environmental interests.

Keywords:

- *truncated Poisson distribution; maximum likelihood estimation; utility function; environment protection.*

AMS Subject Classification:

- 60E05, 60K10, 60N05, 62J05.

✉ Corresponding author.

1. INTRODUCTION

It goes without saying that trees are the main sources of producing papers, until alternatives are proven to be exactly similar, that are utilized for many possible activities in day to day execution of us. Most of the paper mills are in existence for a long time and hence present technologies fall in a wide spectrum ranging from oldest to the most modern. In Indian scenario, the mills use a variety of raw material viz., wood, bamboo, recycled fibre, bagasse, wheat straw, rice husk, etc. In terms of share in total production, approximately 25% are based on wood, 58% on recycled fibre and 17% on agro-residues. India's share in global paper demand is gradually growing as domestic demand is increasing at a steady pace while demand in the western nations is contracting. According to Indian Paper Mill Association, the domestic demand in India grew from 9.3 million tonnes in financial year 2008-09 to 17 million tonnes in financial year 2017-18 at a compound annual growth rate (CAGR) of 6.9%. The futuristic view is that growth in paper consumption would be in multiples of gross domestic product and hence an increase in consumption by one kg per capita would lead to an increase in demand of one million tonnes. Among five important demand drivers, a likely pick-up from the education sector is prominent one. Printing and writing segment demand is expected to grow at a CAGR of 4.2% and reach 5.7 million tonnes in financial year 2020-21 on the back of an anticipated pick-up from the education sector with improving literacy rates and growing enrolment as well as increasing number of schools and colleges.

Therefore, caring nature by reducing the usage of papers is obvious one can easily do, if not striving for a proper alternative that fulfil our need in every possible sense (see Skog and Nicholson [13]; Manzardo *et al.* [8]). Although there are regular plantation of trees required to produce paper products (Rudel [11]), there are several alternatives of non-judicial and unstructured ways of misutilization of the same.

The caring nature in paper usage is an indirect approach of caring by scientifically fulfilling our classical need of papers for examination systems. However, the following two facts are noted in connection to the improper paper utilization in examination systems at the different academic institutions in India. Firstly, students are gradually losing the capacity of writing in case of broad answer type questions, and secondly, the number of pages provided in the main answer scripts during examination are not scientifically matched with actual demand or requirement.

Our objective in the current investigation is three-fold:

(**a**) To identify the distribution of unutilized papers in examination at higher education and to find an optimal setting for number of papers should be provided in an answer script;

(**b**) To find out possible effects due to other variables to the leaving papers blank in examination answer scripts who take the examination in a classical pattern;

(**c**) To address the utilization maximization in view of cost constraints related to answer scripts used.

For the purpose of fulfilling the above objectives, truncated Poisson distribution along with count regression procedure are applied for modelling supported by a real data illustration.

A multiple linear discriminant analysis is also performed in view of grouping into important categories with the help of a real data. The rest of the article is organized as follows. Truncated Poisson is described with its possible applications in section 2. Section 3 deals with count regression models with emphasis given in right truncated Poisson model with mixed effects. In section 4, a real data set on the pages left blank at a semester end examination in a higher education institute in India is been analysed as per the objectives of the research mentioned above and the corresponding results are discussed in dedicated subsections. Section 5 discusses about maximization of a linear utility function of pages in answer scripts subject to certain cost constraints. Finally, the section 6 concludes.

2. TRUNCATED POISSON DISTRIBUTION

The Poisson distribution is a discrete probability distribution usually applied to the number of events occurring within a specified period of time or space. Theoretically, the possible values of a Poisson random variate is non-negative integers (including 0) and there is no upper limit a Poisson random variate can stop for. The Poisson distribution is characterized by a single parameter, usually denoted by $\lambda(> 0)$.

Definition 2.1. A random variable X is said to have a Poisson distribution with parameter λ if its probability mass function (pmf) is of the form

$$\Pr[X = x] = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{for} \quad x = 0, 1, 2, ... \tag{2.1}$$

Numerous applications of Poisson distribution can be found in literature. Some well known applications could be, number of arrivals in a service queue during a specific time interval, number of accidents per month in a city, number of order received per week for a particular product, number of defects in a quality inspection, and number of printing mistakes per page in a book. The wide applicability of Poisson distribution, however, does not lower down its importance, rather newer applications and characterizations are found out in recent years, see Ahmed [1], Johnson *et al.* [5], for more details. Nevertheless, the Poisson distribution is successfully used for situations where some kind of counting is involved.

Truncation in Poisson distribution arises when some specified values are not possible to record (in terms of process and not in terms of availability) either initially or at the end of a Poisson variate range. The former is known as left truncation, while the later is known as right truncation. The theoretical truncated Poisson distribution was introduced by Plackett [10].

Right truncation (omission of values exceeding a specified value r) can occur if the counting mechanism is unable to deal with large numbers or the counting process under consideration is bounded by a finite number.

Definition 2.2. A random variable X is said to have a right truncated Poisson distribution, right truncated at r i.e. the realized values of X is bounded at a specified positive

integer r, with parameter λ if its pmf is of the form

$$\Pr[X = x] = \frac{\lambda^x}{x!}\left(\sum_{j=1}^{r}\frac{\lambda^j}{j!}\right)^{-1}, \qquad x = 0, 1, 2, ..., r. \tag{2.2}$$

If $X_1, X_2, ..., X_n$ are n independent and identically distributed random variables from right truncated Poisson, then the maximum likelihood estimator (MLE) $\hat{\lambda}$ of λ satisfies the following equation:

$$\sum_{j=1}^{r}\frac{(\bar{x} - j)\hat{\lambda}^j}{j!} = 0. \tag{2.3}$$

The simple estimator (Moore [9]) is $\lambda^* = \sum_j \frac{x_j}{m}$, where m is the number of values of x that are less than $r - 1$; this is an unbiased estimator of λ.

3. THE COUNT REGRESSION MODELS

Count data regression models are used for special cases in which the response variable takes count values. It represents the number of events that occur in a given time period. Winkelmann [15] studied the number of live births over a specified age interval of the mother, where the interest was to analyse the variation in terms of the mother's schooling, age, and household income. Another example of count modelling is studied by Cameron *et al.* [2], where they studied the number of times that individuals utilize a health service, such as visits to a doctor or days in the hospital in the past year. The most popular methods to model count data are Poisson and negative binomial regression (Saffari and Adnan [12]). Poisson regression is the more popular of the two and is applied to various fields.

3.1. Poisson regression model

In many situations of practical interest the response variable in an experiment or observational study is a count that is assumed to follow the Poisson distribution. Therefore, a more suitable way to deal with count data is to use the Poisson distribution. The regression model that uses these kinds of option is called the Poisson regression or the Poisson log-linear regression model. For more details use of Poisson regression, one could refer to Frome [4], Lawless [7], Consul and Famoye [3]), Lambert [6] and references therein.

3.2. Truncated Poisson regression models

When the response variable follows a right truncated Poisson distribution, we use right truncated Poisson regression model. In our investigation, to model the number of pages left blank in the main answer booklet in semester end examinations, right truncated Poisson

distribution is utilized owing to the fact that counting is restricted by the total number of available pages in main answer booklet.

There could be three different varieties for right truncated Poisson regression, namely, fixed effect model, random effect model, and mixed effect model. We concentrate in right truncated Poisson regression model for fixed effect on the predictors and random effects for clusters of explanatory variables. Moreover, the random effects to follow a normal distribution with mean 0 and variance σ^2.

3.2.1. Method of estimation

Suppose that we have a sample of n observations $Y_1, Y_2, ..., Y_n$ which can be treated as realizations of independent Poisson random variables, with $Y_i \sim Poi(\lambda_i)$ right truncated at $Y_i \leq r$, and suppose that we want to let the mean λ_i depend on a vector of explanatory variables $\boldsymbol{x_i}$ and random effects. For the Poisson probability function, a model for count data truncated on the right at value r can be expressed as

$$\Pr(Y_i = y_i | Y_i \leq r) = \frac{\Pr(Y_i = y_i)}{\Pr(Y_i \leq r)} = \frac{\lambda_i^{y_i}}{\left(\sum_{k=0}^{r} \frac{\lambda_i^k}{k!}\right) y_i!}, \qquad i = 1, 2, ..., m, \tag{3.1}$$

where m is the number of observation after truncation.

The standard assumption is to use the exponential mean parametrization,

$$\lambda_i = \exp(\boldsymbol{x_i}^T \beta + {z_i}^T u_i), \qquad i = 1, 2, ..., n.$$

In this expression, $\boldsymbol{x_i}$ is a vector of covariates and $\boldsymbol{\beta}$ is a vector of parameters (fixed effect coefficients). The coefficient β can be interpreted as average proportionate change in the conditional mean $E[Y_i|\boldsymbol{x_i}]$ for a unit change is $\boldsymbol{x_i}$. $\boldsymbol{Z}$ is a design matrix of random effects clusters and $\boldsymbol{u}$ is a vector of random effects for that.

In general matrix notation, we can write it as

$$\boldsymbol{\lambda} = \exp(\boldsymbol{X\beta} + \boldsymbol{Zu}), \tag{3.2}$$

where:

$\boldsymbol{X}$: Design matrix of order $n \times p$ for fixed effect explanatory variables;
$\boldsymbol{\beta}$: Vector of fixed effect coefficients;
$\boldsymbol{Z}$: Design matrix of order $n \times q$ for random effect explanatory variables (clusters/groups);
$\boldsymbol{u}$: Vector of random effect coefficients.

The method of hierarchical likelihood method of estimation (h-Likelihood) is used to obtain the values of regression coefficients. Let $Y_{ij} (i = 1, ..., m; j = 1, ..., n_i)$ be the observations of the response variable. Let u_i be the unobserved random effect on the i^{th} individual. We consider the model

$$\Pr(Y_{ij} = y_{ij} | u_i, y_{ij} \leq r) = \frac{\lambda_{ij}^{y_{ij}}}{\left(\sum_{k=0}^{r} \frac{\lambda_{ij}^k}{k!}\right) y_{ij}!} \tag{3.3}$$

such that

$$\lambda_{ij} = \exp({x_{ij}}^T \beta + {z_{ij}}^T u_i)\,, \qquad i = 1, 2, ..., n. \tag{3.4}$$

We assume a normal distribution for the random effects

$$u_i \sim \text{Normal}\big(0, \sigma^2\big). \tag{3.5}$$

Therefore, the h-likelihood (h) is defined by

$$h = L_1(\beta; y|\boldsymbol{u}) + L_2\big(\sigma^2, \boldsymbol{u}\big) \tag{3.6}$$

where $L_1(\beta; y|\boldsymbol{u})$ is the logarithm of the conditional Poisson density function for the response $\boldsymbol{Y}$ given $\boldsymbol{u}$ with parameter $\boldsymbol{\lambda} = \exp(\boldsymbol{X\beta} + \boldsymbol{Zu})$, and $L_2\big(\sigma^2, \boldsymbol{u}\big)$ is the logarithm of the Normal density function for the random effect $\boldsymbol{u}$. Thus,

$$\begin{aligned} L_1(\beta; y|\boldsymbol{u}) &= \sum_{ij}\left[y_{ij}\ln(\lambda_{ij}) - \log(y_{ij}!) - \ln\sum_{k=0}^{r}\frac{\lambda_{ij}^k}{k!}\right] \\ &= \sum_{ij}\left[y_{ij}\big({x_{ij}}^T\beta + {z_{ij}}^T u_i\big) - \ln\sum_{k=0}^{r}\frac{\big(\exp({x_{ij}}^T\beta + {z_{ij}}^T u_i)\big)^k}{k!} - \log(y_{ij}!)\right] \end{aligned} \tag{3.7}$$

and

$$L_2\big(\sigma^2, \boldsymbol{u}\big) = -\sum_i\left[\frac{\ln(2\pi)}{2} + \frac{\ln(\sigma^2)}{2} + \frac{{u_i}^2}{2\sigma^2}\right]. \tag{3.8}$$

The maximum h-likelihood estimators (MHLEs) are obtained by solving the following equations:

$$\frac{\partial h}{\partial \beta_l} = \sum_{ij}\left[y_{ij} - \frac{1}{\sum_{k=0}^{r}\frac{(\exp({x_{ij}}^T\beta + {z_{ij}}^T u_i))^k}{k!}}\sum_k \frac{\big(\exp({x_{ij}}^T\beta + {z_{ij}}^T u_i)\big)^k}{(k-1)!}\right]x_{ijl} = 0$$
$$\text{for } l = 1, ..., p \tag{3.9}$$

and

$$\frac{\partial h}{\partial u_i} = \sum_{j}\left[y_{ij} - \frac{1}{\sum_{k=0}^{r}\frac{(\exp({x_{ij}}^T\beta + {z_{ij}}^T u_i))^k}{k!}}\sum_k \frac{\big(\exp({x_{ij}}^T\beta + {z_{ij}}^T u_i)\big)^k}{(k-1)!}\right]z_{ij} - \frac{u_i}{\sigma^2} = 0$$
$$\text{for } i = 1, ..., m. \tag{3.10}$$

Iterative techniques like, Fisher scoring or Newton-Raphson method can be used to obtain the estimators of the parameters. For more details on the method of estimation for truncated Poisson regression with normal random effects, one could refer to Suaiee [14].

4. APPLICATION WITH REAL LIFE DATA

This section illustrates the methods, described above, with the help of a real data analysis. For the purpose, a sample of 200 students appeared for semester end examination (SEE) are collected from a leading higher education institute in India during November-December, 2018.

Students from various courses and subjects are been considered for balancing possible bias in sampling procedure. However, convenience sampling scheme were applied with adjustments in courses and paper types (quantitative and non-quantitative) for which SEE is taken by the students. Information on the following variables are collected:

1. Course type (under graduate and post graduate);
2. Type of paper written (quantitative and non-quantitative);
3. Number of pages left blank[1];
4. Number of lines written per page[2];
5. Number of words written per line.

For the last three variables, three random observations are taken to ensure unbiasedness and their average is considered.

Statistical software R (version 3.6.0) is utilized for calculations and we see that there are 24% post graduate and 76% undergraduate students in the sample. Quantitative paper was for 56% and non-quantitative for 44%. From Figure 1, we see that the variable pages left blank is normally distributed whereas words written per line is positively skewed. The scatter plots for response variable and predictors are displayed in Figure 2.

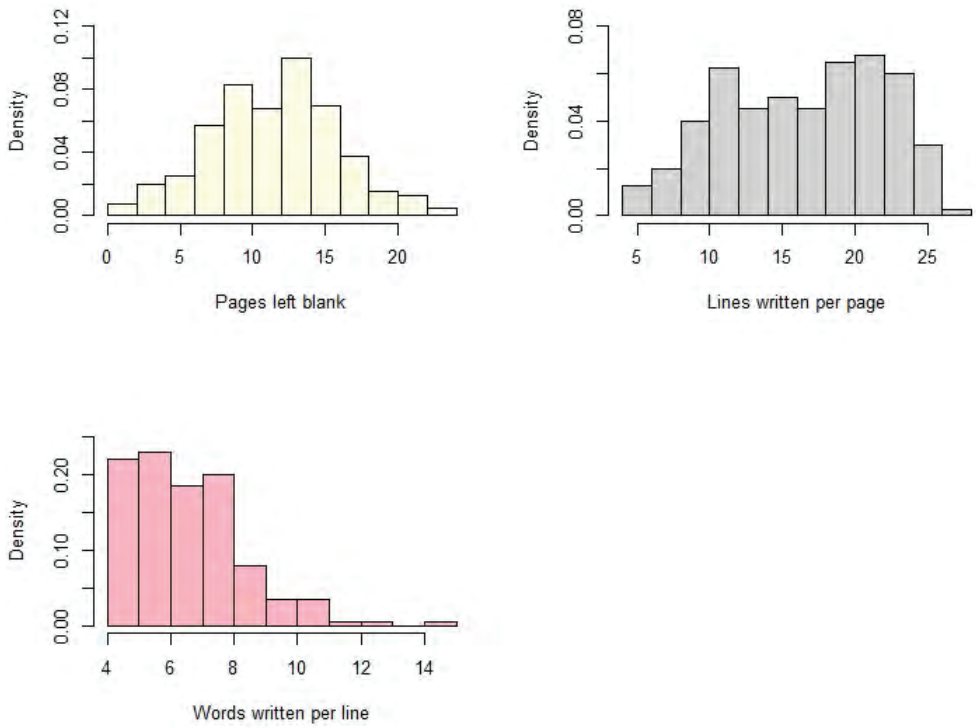
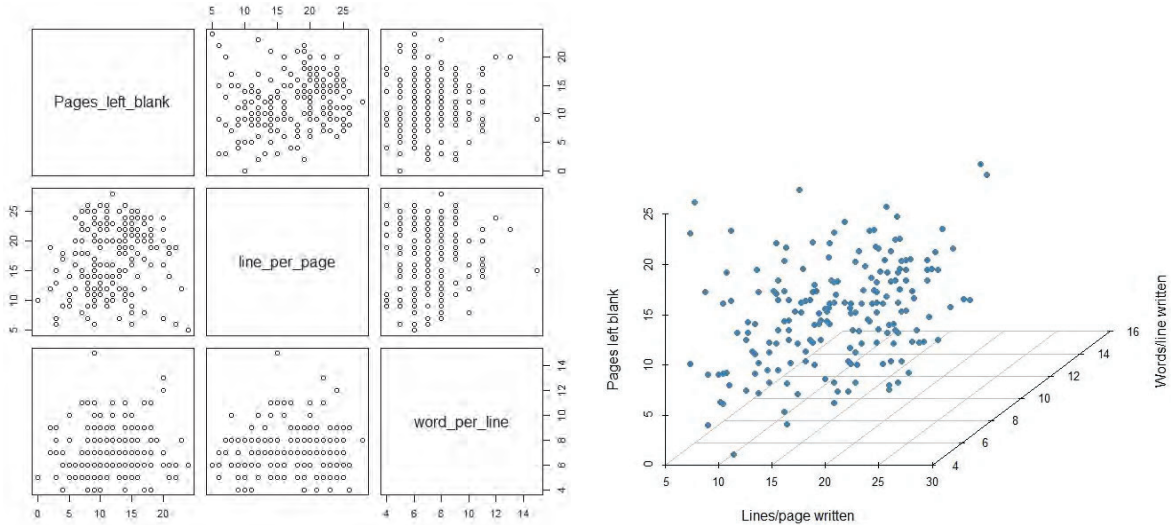


Figure 1: Histograms for response variable and predictors.

[1]The total number of pages in main answer booklet is 25 in the sample collected, excluding front cover page, its immediate back page and one back cover page.

[2]Number of lines per page is 29 in the sample collected.



(a) 2D Scatter plot. **(b)** 3D scatter plot.

Figure 2: Scatter plots for response variable and predictors.

4.1. Justification for using truncated Poisson

Before going to have certain model building on the response variable "pages left blank", let us have the justification for using truncated Poisson distribution (right truncated at 25, the maximum pages in an answer script). We fit the observations on the number of pages left blank with Poisson distribution (without truncation) and right truncated Poisson distributions, respectively. We use maximum likelihood (ML) method of estimation and fitted the Poisson and right Truncated Poisson distributions for the data on the variable "number of pages left blank". As a model selection criteria, the following measures are considered:

(**i**) Akaike information criteria (AIC): $AIC = 2k - 2\ln(\text{loglikelihood})$;

(**ii**) Bayesian information criteria (BIC): $BIC = k\ln(n) + 2\ln(\text{loglikelihood})$.

Here, n: number of observations and k: number of parameters estimated. Lower the values of AIC and BIC, better is the fit. From Table 1, we observe that, right truncated (truncated at 25) Poisson distribution is better for the purpose of modeling. We obtain (refer Table 1) expected number of pages left blank= $11.969 \approx 12$. From Figure 3, we see that the pages left blank data is fitted with right truncated Poisson distribution.

Table 1: ML estimates and model section measures.

Distribution	$\hat{\lambda}$ (Std. Error)	AIC	BIC
Poisson	11.965 (0.24459)	1215.955	1219.253
Right truncated Poisson	11.969 (0.24527)	1214.035	1217.333

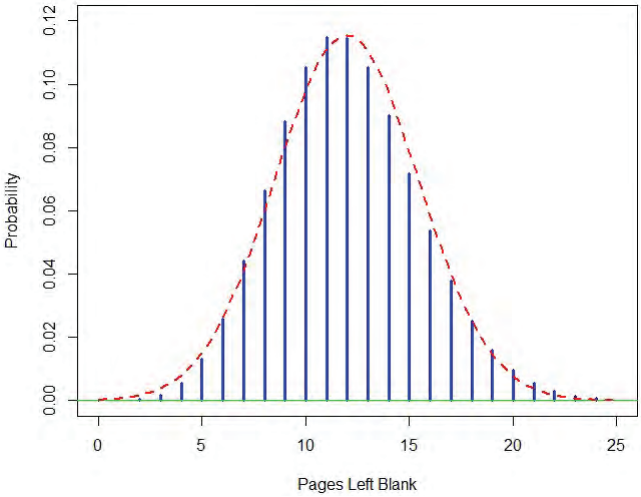


Figure 3: Data fitted with right truncated Poisson and normal curve.

4.2. Truncated Poisson regression with different clusters

In this section we consider response variable as "number of pages left blank". The predictors or explanatory variables are taken as "lines written per page" and "number of words written per line" along with a general mean effect (intercept). We develop three right truncated Poisson regression models considering normal random effects for three different cluster types.

4.2.1. Model-A: Course types as clusters

We consider course type classified as "under-graduate" and "post-graduate" as different clusters having normal random effect. The predictors or explanatory variables are taken as "lines written per page" and "number of words written per line" along with a general mean effect (intercept).

Applying right truncated Poisson regression with normal random effects for course types as clusters, the result obtained is given in Table 2. The log-likelihood, AIC, and BIC values for the model are obtained as -786.5943, 1579.189, and 1589.084, respectively.

Table 2: Regression analysis table: random effects for course type clusters.

Coefficients	Estimate (Std. Error)	t-value	P-value
Intercept	2.03723 (0.10307)	19.766	< 0.0001
Lines per page	0.01666 (0.00378)	4.408	0.00010
Words per line	0.00502 (0.01106)	0.454	0.65000

4.2.2. Model-B: Each individual as cluster

Next, we have considered each individual/student as different clusters having normal random effect. The predictors or explanatory variables are taken as "lines written per page" and "number of words written per line" along with a general mean effect (intercept).

Applying right truncated Poisson regression with normal random effects for individual clusters, the result obtained is given in Table 3. The log-likelihood, AIC, and BIC values for the model are obtained as −711.2469, 1428.494, and 1438.389, respectively.

Table 3: Regression analysis table: random effects for individual clusters.

Coefficients	Estimate (Std. Error)	t-value	P-value
Intercept	2.10554 (0.10326)	20.392	< 0.0001
Lines per page	0.01349 (0.00378)	3.570	0.00036
Words per line	0.00650 (0.01105)	0.588	0.55683

4.2.3. Model-C: Types of paper written as clusters

We next consider type of paper written (classified as quantitative and non-quantitative) as two different clusters having normal random effect. The predictors or explanatory variables are taken as "lines written per page" and "number of words written per line" along with a general mean effect (intercept).

Applying right truncated Poisson regression with normal random effects for clusters, the result obtained is given in Table 4. The log-likelihood, AIC, and BIC values for the model are obtained as −702.2191, 1410.438, and 1420.333, respectively.

Table 4: Regression analysis table: random effects for paper type clusters.

Coefficients	Estimate (Std. Error)	t-value	P-value
Intercept	2.13408 (0.10324)	20.670	<0.0001
Lines per page	0.00768 (0.00377)	2.036	0.0418
Words per line	0.01012 (0.01104)	0.916	0.3595

According to AIC and BIC values, Model-C (types of paper written as clusters) comes out as improved model (refer Table 5). However, for each of the model words written per line is insignificant predictor.

Table 5: Model comparison and information measures.

Coefficients	-Log-likelihood	AIC	BIC
Model-A	-786.5943	1579.189	1589.084
Model-B	-711.2469	1428.494	1438.389
Model-C	-702.2191	1410.438	1420.333

4.3. A linear discrimination approach of grouping

In this section, our objective is to determine whether the variables i.e., pages left blank, lines written per page, and words written per line, will discriminate between quantitative and non-quantitative type paper. Discriminant analysis is a useful multivariate classification technique to predict membership in two or more mutually exclusive groups. We have used paper type (quantitative, non-quantitative) as grouping variable and pages left blank, lines per page, and words per line as independent variables. We have conducted Box's test of homogeneity of covariance matrices and obtained Box's M value as 13.592 which is significant with p-value, $p = 0.038$, to conclude that the groups do differ in their covariance matrices. Wilks' lambda, a measure of how well the discriminant function separates cases into groups, is obtained as 0.543 which is highly significant ($p << 0.05$). The small significance value indicates that the discriminant function does better than chance at separating the groups. The discriminant function is obtained as (considering standardized canonical discriminant function coefficients)

$$D_i = 0.390 \times B_i + 0.923 \times L_i - 0.288 \times W_i, \tag{4.1}$$

where:

D_i: Discriminant score for the i^{th} student;
B_i: Number of pages left blank by the i^{th} student;
L_i: Number of lines written per page by the i^{th} student;
W_i: Number of words written per line by the i^{th} student.

The cut-off value of discriminant score is calculated by taking average of group centroids (mean discriminant score for each group) and is obtained as 18.79. The model will classify any paper as quantitative if the discriminant score is less than 18.79 and non-quantitative otherwise. For example, if we take a random observation i.e. an answer script having 9 pages left blank, 22 lines written per page and 7 words written per line; the discriminant score is obtained as $0.390 \times 9 + 0.923 \times 22 - 0.288 \times 7 = 21.8$, which means this answer script would be classified as a non-quantitative paper type. The classification result (i.e. actual versus predicted group membership) is shown in Table 6, where the overall 81.5% actual group cases are correctly classified.

Table 6: Classification results.

Actual membership	Predicted membership	
	Quantitative	Non-quantitative
Quantitative	86 (72%)	26 (23%)
Non-quantitative	12 (14%)	76 (86%)

The following important findings along with specific recommendations are noted in this section:

1. The expected number of pages left blank in main answer script is 12, i.e., expected number of pages written is 13. We recommend to utilize the residual pages that are not used in main answer scripts for producing additional answer sheets (each with 4 pages composition). The benefit in doing so is that there could be a reduction in making cost and wastage of pages would be minimized as additional sheets can be used whenever required.
2. Types of paper written came out as an important predictor for the response variable, pages left blank, and hence is a meaningful grouping in discrimination.
3. A cut off score of 18.79 discriminates an answer script in two non-overlapping categories if certain minimal information is provided.

This next section discusses about a possible maximization of utility of pages in a single semester of any particular year.

5. DISCUSSION ON UTILITY MAXIMIZATION

The main objective in this section is to discuss about a maximization aspect of the difference of page utility from current to the modified page numbers, subject to costs incurred for such modification and to identify the optimal reduction required in number of pages in answer script. An utility maximization problem can be framed as below:

We define:

X: Number of pages currently used in a main answer script;
N_1: Number of main answer script used in any examination;
c_{11}: Making cost per page for an answer script with X number of pages;
X_1: Number of pages should be used (after reduction following the procedure described in section 4) in a main answer script;
c_{12}: Making cost per page for an answer script with X_1 number of pages;
c_{21}: Making cost per page for an additional answer script with 4 number of pages;
N_2: Number of additional answer script used in any examination;
c_{22}: Per unit cost for making additional $(X - X_1)/4$ number of additional answer script.

Assuming a linear function, let us now define the current and revised utility in terms of total pages that can be utilized in the whole process:

Current utility: $(N_1X + 4N_2)$.
Revised utility: $N_1X_1 + 4N_2 + \frac{N_1(X-X_1)}{4}$.

We want to

$$\begin{aligned} \text{Maximize } U(X, X_1) &= (N_1X + 4N_2) - \left[N_1X_1 + 4N_2 + \frac{N_1(X - X_1)}{4}\right] + k \\ &= \frac{3N_1(X - X_1)}{4} + k, \end{aligned} \tag{5.1}$$

where k is an integer constant and $N_1(X - X_1) \equiv k \,(\mathrm{mod}\, 4)$.

Subject to the constraints,

$$N_1 X c_{11} - N_1 X_1 c_{12} \geq A_0 \tag{5.2}$$

(surplus cost inequation for main answer script),

$$4N_2 c_{21} + \frac{N_1(X - X_1)}{4} c_{22} \leq A_0 \tag{5.3}$$

(cost inequation for additional answer script),

$$\text{with } X, X_1 \geq 0. \tag{5.4}$$

Here, A_0 is amount of threshold benefit which is known or specified.

Now, for given values of $c'_{ij}s; i = 1, 2, j = 1, 2$ and known N_1, N_2, one can easily optimize (an integer programming problem) the function in (5.1) for X and X_1.

6. CONCLUDING REMARKS

This article provides a scientific way of allocating pages in main answer scripts in classical examination system in higher education sector. The study is restricted to one particular higher education institute in India. However, scope for investigations are open for multi-centric observations in the different educational institute in the same country and/or foreign institutes. Estimate for the number of pages blank will be an important investigation for multi-centric study as all the higher education institutes do not provide same number of pages in main answer scripts. We hope this article shall provide the authorities, all stake holders, and the student community, an alarming consciousness about the proper utilization of the pages used for education and thereby shall protect the environment thinking the large scale impact of the same to the environment.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the anonymous referees for their valuable comments and suggestions that have improved the manuscript in its revised version.

REFERENCES

[1] Ahmed, A.H. (1991). Characterization of beta, binomial, and Poisson distributions, *IEEE Transactions On Reliability*, **40**(3), 290–295.

[2] Cameron, A.C.; Trivedi, P.K.; Milne, F. and Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in Australia, *The Review of Economic Studies*, **55**(1), 85–106.

[3] Consul, P. and Famoye, F. (1992). Generalized Poisson regression model, *Communications In Statistics – Theory And Methods*, **21**(1), 89–109.

[4] Frome, E.L. (1983). The analysis of rates using Poisson regression models, *Biometrics*, **39**(3), 665–674.

[5] Johnson, N.L.; Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions*, Vol. 444, John Wiley & Sons.

[6] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, **34**(1), 1–14.

[7] Lawless, J.F. (1987). Negative binomial and mixed Poisson regression, *Canadian Journal of Statistics*, **15**(3), 209–225.

[8] Manzardo, A.; Ren, J.; Piantella, A.; Mazzi, A.; Fedele, A. and Scipioni, A. (2014). Integration of water footprint accounting and costs for optimal chemical pulp supply mix in paper industry, *Journal of Cleaner Production*, **72**, 167–173.

[9] Moore, P.G. (1954). A note on truncated Poisson distributions, *Biometrics*, **10**(3), 402–406.

[10] Plackett, R.L. (1953). The truncated Poisson distribution, *Biometrics*, **9**(4), 485–488.

[11] Rudel, T.K. (2009). Tree farms: driving forces and regional patterns in the global expansion of forest plantations, *Land Use Policy*, **26**(3), 545–550.

[12] Saffari, S.E. and Adnan, R. (2011). Zero-inflated Poisson regression models with right censored count data, *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, **27**, 21–29.

[13] Skog, K.E. and Nicholson, G.A. (1998). Carbon cycling through wood products: the role of wood and paper products in carbon sequestration, *Forest Products Journal*, **48**(7), 75–83.

[14] Suaiee, A.M.A. (2013). *Double Truncated Poisson Regression Model with Random Effects*, Ph.D Thesis, University of North Colorado.

[15] Winkelmann, R. (1995). Duration dependence and dispersion in count-data models, *Journal of Business & Economic Statistics*, **13**(4), 467–474.

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 77–96
https://doi.org/10.57805/revstat.v21i1.397

A Multivariate Quantile Based on Kendall Ordering

Authors: Matthieu Garcin ✉
– Léonard de Vinci Pôle Universitaire, Research Center,
92 916 Paris La Défense, France
matthieu.garcin@m4x.org

Dominique Guégan
– Université Paris 1 Panthéon-Sorbonne, LabEx ReFi,
and Ca'Foscari University of Venezia,
106 bd de l'Hôpital, 75013 Paris, France
dguegan@univ-paris1.fr

Bertrand Hassani
– Université Paris 1 Panthéon-Sorbonne, LabEx ReFi,
and University College London – Computer Science,
106 bd de l'Hôpital, 75013 Paris, France
bertrand.hassani@gmail.com

Received: March 2020 Revised: September 2021 Accepted: September 2021

Abstract:

- We introduce the Kendall multivariate quantiles, which are a transformation of orthant quantiles by the Kendall function. Each quantile is then a set of vectors with some advantageous properties, compared to the standard orthant quantile:
  - **i**) it induces a total order
  - **ii**) the probability level of the quantile is consistent with the probability measure of the set of the dominated vectors,
  - **iii**) the multivariate quantiles based on the distribution function or on the survival function have vectors in common which conciliate both upper- and lower-orthant approaches.

  Definition and properties of the Kendall multivariate quantiles are illustrated using Archimedean copulas.

Keywords:

- *multivariate quantile; copula; Archimedean copula; Kendall distribution; orthant quantile.*

AMS Subject Classification:

- 91G70, 62H05, 62P20.

✉ Corresponding author.

1. INTRODUCTION

Given a random variable X defined on a probability space, the quantile of X at level α, Ψ_α, is such that $\alpha = \mathbb{P}[X \leq \Psi_\alpha]$. However, while univariate quantiles are well documented [3], the multivariate approach is not as straightforward. The multivariate analysis relies on a multivariate probability distribution. A useful tool for representing multivariate distributions is the copula which describes the dependence between random variables. A copula is a d-dimensional distribution function $[0,1]^d \to [0,1]$, where $d \in \mathbb{N}$ is the number of underlying random variables [31]. Let G be the multivariate probability distribution of a random vector $X = (X_1, ..., X_d)'$, i.e. the copula-based probability distribution:

$$G : (y_1, ..., y_d) \in \mathbb{R}^d \mapsto \mathbb{P}[X_1 \leq y_1, ..., X_k \leq y_d].$$

If $G_1, ..., G_d$ are the d univariate marginal distribution functions of X, then Sklar's theorem affirms the existence of a copula C such that $G(y_1, ..., y_d) = C(G_1(y_1), ..., G_d(y_d))$ [40]. Copulas, as a tool describing the dependence between random variables, have been applied in many fields, mostly in finance, but also in hydrology [20], astronomy [37, 36], or in telecommunication networks [33, 19].

The multivariate quantile of X can be defined as the set of vectors belonging to the boundary of the α-level set of G. In this set-valued approach, we distinguish the lower-orthant quantile, $\underline{\Psi}_\alpha(G)$, and the upper-orthant quantile, $\overline{\Psi}_\alpha(G)$. The lower-orthant quantile is defined by the set of vectors

$$\underline{\Psi}_\alpha(G) = \partial\{y \in \mathbb{R}^d | G(y) \geq \alpha\}, \tag{1.1}$$

where ∂ denotes the boundary of the mentioned set, whereas the upper-orthant quantile is

$$\overline{\Psi}_\alpha(G) = \partial\{y \in \mathbb{R}^d | \overline{G}(y) \leq 1 - \alpha\}, \tag{1.2}$$

where $\overline{G}$ is the survival function associated with G. If $(y_1, ..., y_d)' \in \underline{\Psi}_\alpha(G)$, then

$$\mathbb{P}[X_1 \leq y_1, ..., X_d \leq y_d] = \alpha.$$

In general, the lower-orthant quantile is more conservative than the upper-orthant quantile: this means that if a vector y belongs to both set-valued orthant quantiles $\overline{\Psi}_\alpha(G)$ and $\underline{\Psi}_{\alpha'}(G)$, then the probability level α' associated by the lower-orthant quantile to y is lower than the probability level α associated by the upper-orthant quantile to y.

The notion of multivariate quantile has been studied in various ways in the literature [38]. For instance we can mention the approach by Embrechts and Puccetti's [16] based on the orthant quantiles defined in equations (1.1) and (1.2) for applications in finance and insurance [11]. Outside the field of finance, the multivariate quantiles have been also studied [38], and applied in particular to meteorology, where extreme weather depends on a combination of parameters which cannot be aggregated, such as speed of wind, quantity of precipitation, temperature, and cloud cover [28] or in hydrology for frequency analysis [9]. These fields require advanced methodological and theoretical support with respect to multivariate analysis. Indeed, when dealing with multivariate data, no consensus arises about the definition of order statistics and quantiles. In particular, the question of quantiles of multivariate distributions has led to numerous interpretations often inspired by analogies with

different ways of defining the quantiles of a univariate distribution. Among the various methods proposed, we can cite the spatial quantile [1, 39, 7, 14] or the geometric quantile [8, 6], with some applications in finance [24]. The inversion of a mapping is another kind of known multivariate quantile. In the unidimensional framework, a quantile is indeed defined as the generalized inverse of the cumulated distribution function. If one defines a mapping F from $\mathbb{R}^d$ to $\mathbb{R}$, then inversions can also define a quantile [27]. The exact definition of a multivariate quantile based on the inversion of a mapping is provided by equation (1.1), where the distribution function G is to be replaced by the mapping F. This method is linked to multivariate ordering based on a scalarization, which is the ordering of vectors by comparing scalars, such as isolated coordinates or a function of a linear combination of coordinates [34], or any mapping [41]. This is the method used for example in the orthant quantile, with F being in this case the joint distribution of the d coordinates.

Though extensions of orthant quantiles have been proposed, for instance by Cousin and Di Bernardino [12], who replaced both sets defined by the lower-orthant and the upper-orthant quantiles by their expected value, engendering a quantile defined by a simple vector instead of an infinite set of vectors. Also, replacing the set-valued orthant quantile by a vector-valued quantile has been made possible by selecting a particular direction [42]. The vector-valued multivariate quantile is then the intersection of the set-valued quantile with a line in $\mathbb{R}^d$, given the arbitrary choice of the direction of this line. These two singularizations of the orthant quantile show a need to be able to compare and order multivariate quantiles of different confidence levels. But a pitfall of the orthant approach is that it does not induce a total order, as defined below.

Definition 1.1. Given a random vector X defined in a probability space, we consider a set-valued multivariate quantile function $\alpha \in [0,1] \mapsto \Psi_\alpha(F)$ based on the inversion of a mapping $F : \mathbb{R}^d \to \mathbb{R}$ defined as the probability of a subset of $\mathbb{R}^d$. In other words, $\forall y \in \mathbb{R}^d$, $\exists\, \mathcal{S}_y \subset \mathbb{R}^d, F(y) = \mathbb{P}[X \in \mathcal{S}_y]$. We provide the following definitions:

- The vector $y \in \mathbb{R}^d$ is said to dominate the vector $z \in \mathbb{R}^d$ if $z \in \mathcal{S}_y$. We write it $z \preceq y$. If $y \in \Psi_\alpha(F)$, α is the probability of the set of all the vectors dominated by y.
- The order induced by this set-valued quantile is said to be total if $\preceq$ is a total order. In particular, $\forall y, z \in \mathbb{R}^d$, we have in this case $y \preceq z$ or $z \preceq y$. If this property does not hold, the order is said to be partial.

For the lower-orthant quantile, $\mathcal{S}_y$ is simply the lower-left orthant of y, that is the set of vectors for which each component is lower than the corresponding component of y. It could be interesting to extend the orthant approach, which induces a partial order, to a total order. We think that this total order is a desirable property for a multivariate set-valued quantile. Indeed, we consider that if y and z are vectors of $\Psi_\alpha(F)$ they should dominate the same set of vectors, this property leading to a total order. Furthermore, in this case, every vector $x \in \Psi_{\alpha'}(F)$, with $\alpha' < \alpha$, is dominated by y and z. This property does not hold for instance in the orthant approach. The direct consequence of this property is that α is solution of the equation

$$\alpha = \mathbb{P}[X \in \Psi_{\alpha'}(F) | \alpha' < \alpha]. \tag{1.3}$$

In other words, the probability measure of the set-valued quantiles $\Psi_{\alpha'}(F)$ for a probability level α' lower than α is exactly α, similarly to the univariate case. To our knowledge, existing set-valued multivariate quantiles, including orthant quantiles, do not fulfill this property.

In the family of multivariate quantiles based on the inversion of a mapping F, a proper choice of F may lead to a total order. We are interested in finding this proper F. In this quest, we are inspired by another setting of multivariate quantile known as centre-outward quantile surface. If one is given a statistical depth function, such as the likelihood depth [18] or the Mahalanobis depth [30], the centre-outward quantile surfaces are defined as concentric regions around the centre, which is the maximal-depth vector [29, 43]. More precisely, given a probability $p \in (0,1)$, the p-quantile of a distribution G is the set of vectors of depth α_p, which is defined such that the probability to have vectors of a higher depth than α_p is p: $p = \mathbb{P}(D(X,G) \geq \alpha_p)$, where X is a random vector of distribution G and $D(X,G)$ its depth.

The purpose of this paper is to propose an extension of the orthant quantile that also induces a total order and in particular for which equation (1.3) holds, and to study its properties. We propose to modify the centre-outward quantile surface to focus on tails instead of on the centre of the distribution. Instead of determining a spatial median first, we associate a metric for each vector. Vectors with a metric of the same value are gathered in an equivalence class. We can then order these classes with respect to this metric. The metric chosen is the multivariate distribution function and is thus consistent with the orthant quantile approach. In the quantile surface approach, if y belongs to the p-quantile, then the metric associated to any random vector X is lower with probability p than the one associated with y. For this reason, we will use the Kendall probability distribution

$$K : t \in [0,1] \mapsto \mathbb{P}[G(X) \leq t],$$

where G is the multivariate probability distribution function of X, applied to the the random vector X. The Kendall function indeed defines natural equivalence classes [32]. If vectors y and z are such that $G(y) = G(z)$, then the vector y is equivalent to z, and these vectors dominate every vector x such that $G(x) < G(y)$. Contrary to the orthant quantile, we affirm that y is a vector belonging to the set of the quantile of probability $K(G(y))$, instead of a probability $G(y)$, which is lower than $K(G(y))$ by construction. We base our new definition on the Kendall stochastic ordering [32] instead of the traditional product ordering. The first one is a total order, whereas the second one is only partial. An explanatory illustration is provided in Figure 1.

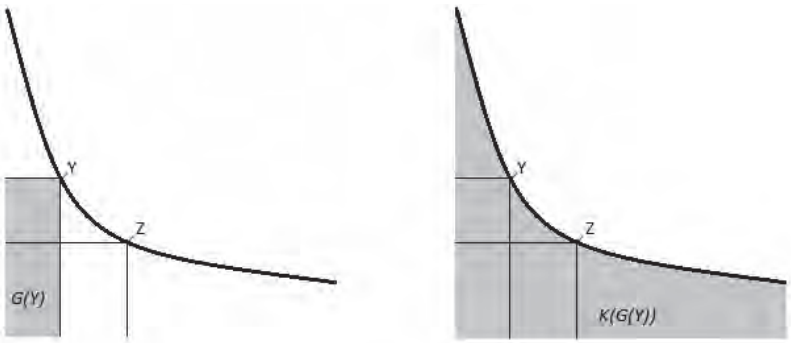


Figure 1: On the left, the thick line is a set of two-dimensional vectors having the same lower-left cumulated probability. In particular, $G(y) = G(z)$, which is the probability measure of the lower-left quadrant of y or z. However, some vectors of the lower-left quadrant of z are not in the lower-left quadrant of y and therefore cannot be compared to y in terms of dominance. On the right, the multivariate probability distribution only leads to the definition of equivalence classes. Therefore, every vector in the grey zone is dominated by every vector on the thick line. The vectors dominated by z are the same as those dominated by y. The probability associated with y and z is therefore the probability measure of the entire grey zone, that is $K(G(y))$, which is equal to $K(G(z))$ and which is greater than $G(y)$.

This blend of orthant quantile and quantile surface leads to a new definition of the multivariate quantile. We call it the lower-orthant or upper-orthant Kendall multivariate quantile, since it uses the Kendall distribution function. The lower-orthant Kendall multivariate quantile is $\underline{\Psi}_{\alpha}^{K}(G) = \partial\{y \in \mathbb{R}^d | K(G(y)) \geq \alpha\}$. The upper-orthant Kendall multivariate quantile is $\overline{\Psi}_{\alpha}^{\mathcal{K}}(G) = \partial\{y \in \mathbb{R}^d | \mathcal{K}(\overline{G}(y)) \leq 1-\alpha\}$, where $\mathcal{K} : t \in [0,1] \mapsto \mathbb{P}[\overline{G}(X) \leq t]$. According to Definition 1.1, the Kendall multivariate quantile induces a total order, contrary to the orthant quantile.

In this paper, we present some properties of this multivariate quantile. For instance, we will observe the extent to which the Kendall multivariate quantiles differ from the orthant quantiles. In particular, the Kendall multivariate quantile is less conservative than the lower-orthant quantile and more conservative than the upper-orthant quantile. Indeed, if a vector y belongs to $\underline{\Psi}_{\underline{\alpha}}(G)$, $\overline{\Psi}_{\overline{\alpha}}(G)$, $\underline{\Psi}_{\underline{\alpha}^K}^{K}(G)$, and $\overline{\Psi}_{\overline{\alpha}^{\mathcal{K}}}^{\mathcal{K}}(G)$, then the probability levels associated to this vector are ordered in the following way: $\underline{\alpha} \leq \underline{\alpha}^K \leq \overline{\alpha}$ as well as $\underline{\alpha} \leq \overline{\alpha}^{\mathcal{K}} \leq \overline{\alpha}$. Moreover, nothing indicates which of the lower-orthant quantile and of the upper-orthant quantile should be preferred. The Kendall quantile can then be seen as a way of diminishing the impact of such a choice, because there is also a smaller difference between both Kendall quantiles than between both orthant quantiles: $|\overline{\alpha}^{\mathcal{K}} - \underline{\alpha}^K| \leq |\overline{\alpha} - \underline{\alpha}|$.

In Section 2, we introduce the Kendall multivariate quantile and some of its properties. We provide theoretical results comparing the Kendall multivariate quantile with the orthant ones. In Section 3, we focus on the case of Archimedean copulas and present an application to simulated data. Section 4 concludes our findings.

2. KENDALL'S MULTIVARIATE QUANTILE

Two approaches relying on Kendall distributions are presented in the next subsection. Drawing a parallel with lower- and upper-orthant multivariate quantiles presented above, we formalise the notion of lower- and upper-orthant Kendall quantile. But before this, we state an assumption that will hold in the whole article.

Assumption 2.1. All the copulas considered have no singular components.

2.1. Definitions

1. *Lower-orthant Kendall quantile.*

As introduced in the section above, the Kendall distribution function is $K : t \in [0,1] \mapsto \mathbb{P}[G(X) \leq t]$, where G is the multivariate probability distribution of the random vector X, associated with a given copula. It is worth noting that the Kendall function does not depend on the full distribution of X but only on its dependence structure. The Kendall function has been used, for example, to estimate Archimedean copulas [21] or to create hierarchical Kendall copulas that deal with high-dimension problems [4]. Using this function, we define the lower-orthant Kendall quantile.

Definition 2.1. For a random vector X of dimension d, the lower-orthant Kendall quantile of probability $\alpha \in [0,1]$, denoted $\underline{\Psi}_\alpha^K$, is the boundary set of the set of vectors $y \in \mathbb{R}^d$ such that $K(G(y)) \geq \alpha$, where G is the multivariate distribution of X and K the corresponding Kendall function:

$$\underline{\Psi}_\alpha^K(G) = \partial\{y \in \mathbb{R}^d | K(G(y)) \geq \alpha\}.$$

2. *Upper-orthant Kendall quantile.*

Similar to the distinction between lower-orthant and upper-orthant multivariate quantiles, we can make a distinction between two kinds of Kendall quantiles, based either on the multivariate distribution function G or on the corresponding survival function $\overline{G}$. We thus introduce another Kendall function in Definition 2.2, $\mathcal{K} : t \in [0,1] \mapsto \mathbb{P}[\overline{G}(X) \leq t]$. We stress the fact that $\mathcal{K}$ is neither the survival Kendall function associated to G, nor K itself, but it is the standard Kendall function associated to $\overline{G}$.

Definition 2.2. For a random vector X of dimension d, the upper-orthant Kendall quantile of probability $\alpha \in [0,1]$, denoted $\overline{\Psi}_\alpha^{\mathcal{K}}$, is the boundary set of the set of vectors $y \in \mathbb{R}^d$ such that $\mathcal{K}(\overline{G}(y)) \leq 1-\alpha$, where $\overline{G}$ is the survival function associated to the multivariate distribution G of X, and $\mathcal{K}$ is the Kendall function corresponding to $\overline{G}$, that is $\mathcal{K} : t \in [0,1] \mapsto \mathbb{P}[\overline{G}(X) \leq t]$:

$$\overline{\Psi}_\alpha^{\mathcal{K}}(G) = \partial\{y \in \mathbb{R}^d | \mathcal{K}(\overline{G}(y)) \leq 1-\alpha\}.$$

These different definitions of multivariate quantiles are linked, as exposed in the following proposition. Indeed, contrary to lower-orthant and upper-orthant multivariate quantiles, both Kendall's multivariate quantiles have some vectors in common.

Proposition 2.1. *Let $\alpha \in [0,1]$, G be a non-atomic multivariate distribution function of dimension $d \in \mathbb{N}$, and having a density function whose support is $\mathbb{R}^d$, with K the Kendall function, supposed to be strictly monotonic, $\overline{G}$ the survival distribution, both associated with G, $\mathcal{K}$ the Kendall function of $\overline{G}$ as introduced in Definition 2.2, then:*

$$\overline{\Psi}_\alpha^{\mathcal{K}}(G) \cap \underline{\Psi}_\alpha^K(G) \neq \emptyset.$$

The proof of Proposition 2.1 is reported in the Appendix.

2.2. Properties

In this section, we focus on specific properties of the Kendall multivariate quantile. In particular, we specify the difference between the Kendall quantile and the orthant quantile.

As mentioned above, the probability associated with a vector by the Kendall function is higher (respectively lower) than in the lower-orthant (resp. upper-orthant) approach.

Using the Fréchet-Hoeffding bounds, it can be demonstrated that $t \leq K(t) \leq 1$ [21] and that $t \leq \mathcal{K}(t) \leq 1$ as well. For a vector y and a multivariate probability distribution G, the lower-orthant approach links y to the level of probability $G(y)$, whereas the Kendall approach associates it with a probability $K(G(y))$, which is, therefore, in the interval $[G(y), 1]$. In other words, the two approaches provides a different probability for a same vector. This probability is higher in the lower-orthant Kendall approach than in the lower-orthant one. We can compare both quantiles in the following manner:

Proposition 2.2. *Let K be strictly monotonic on a neighbourhood of a given probability $\alpha \in [0, 1]$. Then, the Kendall quantile and the lower-orthant quantile are linked by the following:*

$$\underline{\Psi}_\alpha(G) = \underline{\Psi}^K_{K(\alpha)}(G).$$

The proof of Proposition 2.2 is reported in the Appendix.

Similarly, we can show that the upper-orthant Kendall quantile (which has a non-empty intersection with the lower-orthant Kendall quantile as stated in Proposition 2.1) associates a vector with a lower probability than does the upper-orthant quantile. It is the meaning of the next proposition, since $\mathcal{K}$ is a growing function and since we have $1 - \mathcal{K}(1 - \alpha) \leq \alpha$.

Proposition 2.3. *Let $\mathcal{K}$ be strictly monotonic on a neighbourhood of a given probability $\alpha \in [0, 1]$. Then, the upper-orthant Kendall quantile and the upper-orthant quantile are linked by the following:*

$$\overline{\Psi}_\alpha(G) = \overline{\Psi}^{\mathcal{K}}_{1-\mathcal{K}(1-\alpha)}(G).$$

The proof is similar to the one of Proposition 2.2 and is thus omitted. We can also compare the level associated with the lower-orthant Kendall quantile to the level associated to the upper-orthant quantile, and the comparison can also be between the upper-orthant Kendall quantile and the lower-orthant quantile.

Proposition 2.4. *Let $\alpha, \alpha', \alpha'' \in [0, 1]$ and G be a probability distribution with no atoms.*

1. *If $\underline{\Psi}_\alpha(G) \cap \overline{\Psi}^{\mathcal{K}}_{\alpha''}(G) \neq \emptyset$, then $\alpha'' \geq \alpha$.*
2. *If $\overline{\Psi}_{\alpha'}(G) \cap \underline{\Psi}^K_{\alpha''}(G) \neq \emptyset$, then $\alpha'' \leq \alpha'$.*

The proof of Proposition 2.4 is reported in the Appendix.

The message conveyed by Propositions 2.2, 2.3, and 2.4 is that both the Kendal quantiles are a compromise between both orthant quantiles.

An interesting metric to compare the lower-orthant Kendall quantile and the lower-orthant quantile is given by the positive function $r : \alpha \in [0, 1] \mapsto K(\alpha) - \alpha$. This function r is the difference of probability associated with a same vector by the lower-orthant Kendall quantile and by the lower-orthant quantile, for a given level of probability. In other words, for a probability α, $\underline{\Psi}_\alpha(G)$ is a set of vectors corresponding to this probability α. For the same set of vectors, the lower-orthant Kendall quantile associates another level of probability, which is $K(\alpha)$ according to Proposition 2.2, and $r(\alpha)$ denotes this difference of probabilities.[1]

[1] For example, the Gumbel copula in Example 3.1 leads to $r(\alpha) = -\frac{\alpha \log(\alpha)}{\theta}$.

Generally, r can be linked to the Kendall rank correlation coefficient, known as Kendall's tau coefficient, as stated by the following proposition.

Proposition 2.5. *The average difference between the probabilities associated to the Kendall function and to the sole copula, for d-dimensional vectors and a continuous copula, is the following:*

$$\int_0^1 r(\alpha)d\alpha = (1-\tau)\left(\frac{1}{2} - \frac{1}{2^d}\right),$$

where τ is the Kendall rank correlation coefficient.

The proof of Proposition 2.5 is reported in the Appendix.

In the bivariate case, this average difference is $(1-\tau)/4$ which belongs to $[0, 1/2]$, due to the fact that $\tau \in [-1, 1]$. When d tends toward infinity, the average difference increases concomitantly with the dimension d, up to $(1-\tau)/2 \in [0, 1]$. The case of the independent copula, for which $\tau = 0$, leads to an average r of $(1/2) - (1/2)^d$, whose value, $1/4$ for $d = 2$, progressively increases with the dimension up to $1/2$. It confirms the analysis presented in Example 3.2. If we consider comonotonic coordinates, then $\tau = 1$ and the average r is equal to zero, whatever the dimension d. Graphically, it corresponds to a case where all the vectors dominated by a reference vector belong to the lower-left quadrant of this reference vector. The order implied by the orthant quantiles, which is partial in general, is total in this particular case, and there is no difference between the orthant and the Kendall quantiles. In the case of the opposite, if the coordinates are countermonotonic then $\tau = -1$ and the average r reaches its maximum, $1 - (1/2)^{d-1}$, which goes from $1/2$, for $d = 2$, to 1, when d goes toward infinity.

Additionally, we can quantify the difference between the probability associated to a vector by the upper-orthant method and by the upper-orthant Kendall method: $\overline{r} : \alpha \in [0, 1] \mapsto \alpha - (1 - \mathcal{K}(1-\alpha))$ which is a positive function. Proposition 2.6 states that the average twist $\overline{r}$ of the probability level between the upper-orthant quantile and the upper-orthant Kendall quantile is, in absolute value, exactly the same as the average twist r between the lower-orthant quantile and the standard Kendall quantile.

Proposition 2.6. *The average difference between the probabilities associated to the sole survival copula and to the Kendall function of the survival copula, for d-dimensional vectors and a continuous copula, is the following:*

$$\int_0^1 \overline{r}(\alpha)d\alpha = (1-\tau)\left(\frac{1}{2} - \frac{1}{2^d}\right),$$

where τ is the Kendall rank correlation coefficient.

The proof of Proposition 2.6 is reported in the Appendix.

In the framework of Proposition 2.1, where the upper-orthant and the lower-orthant Kendall quantiles have a non-empty intersection for a given probability level, the vectors belonging to both Kendall quantiles can thus be seen as a balanced compromise between lower- and upper-orthant quantiles. Indeed, in absolute value, they twist the probability associated with both in average over all the possible probability levels similarly, as stated in

Propositions 2.5 and 2.6. Nevertheless, for a particular level of probability, the lower-orthant Kendall quantile can be closer to one or to the other.

In addition to the average error, we can calculate an upper bound of the limit error $r(\alpha)$, when α tends to 0 or 1, and more widely of the probability distortion between the lower-orthant and the upper-orthant quantiles. We call probability distortion the difference of probability according to the lower-orthant and the upper-orthant approach for a vector x belonging to both sets: if $x \in \underline{\Psi}_\alpha(G) \cap \overline{\Psi}_{\alpha'}(G)$, the probability distortion is $\alpha' - \alpha$. It depends on α but also on the choice of x in $\underline{\Psi}_\alpha(G)$. To establish ideas, we will focus on a particular x corresponding to equal marginal probabilities: $x = (G_1^{-1}(u), ..., G_d^{-1}(u))$, where $u \in [0,1]$ is well chosen to have $x \in \underline{\Psi}_\alpha(G)$. So we have $u = \delta^{-1}(\alpha)$, where $\delta : v \in [0,1] \mapsto C(v, ..., v)$ is the diagonal section of the copula C associated to the joint distribution G. This choice is possible only if δ is invertible. As x is an element of the set $\underline{\Psi}_\alpha(G)$, the probability associated with x in the lower-orthant approach is α. By definition of the upper-orthant quantile, x is also an element of the set $\overline{\Psi}_{\alpha'}(G)$ with $\alpha' = 1 - \overline{G}(x) = 1 - \overline{G}\big(G_1^{-1}\big(\delta^{-1}(\alpha)\big), ..., G_d^{-1}\big(\delta^{-1}(\alpha)\big)\big)$. If we note $\alpha \mapsto R(\alpha)$ the function of distortion of probability between the lower-orthant and the upper-orthant quantiles, then $R(\alpha) = \alpha' - \alpha$, which we can equivalently write:

$$R : \alpha \in [0,1] \mapsto 1 - \alpha - \overline{G}\big(G_1^{-1}\big(\delta^{-1}(\alpha)\big), ..., G_d^{-1}\big(\delta^{-1}(\alpha)\big)\big). \tag{2.1}$$

In Proposition 2.7, we propose an upper bound for $R(\alpha)/\alpha$.

This distortion $R(\alpha)$ is directly linked to the notion of tail dependence. For a bivariate variable, the lower tail dependence λ_L is the following limit, if it exists: $\lim_{\alpha \to 0} \mathbb{P}(X_1 \leq G_1^{-1}(\alpha) | X_2 \leq G_2^{-1}(\alpha))$. Owing to Bayes' rule, this expression is symmetric in each component of the vector. Moreover, it only depends on the copula and not on the marginals. In higher dimension, one can define several lower tail dependence parameters corresponding to various choices of subsets $I_k \subset \{1, ..., d\}$ of size k: $\lambda_{L,I_k} = \lim_{\alpha \to 0} \mathbb{P}(X_i \leq G_i^{-1}(\alpha), \forall i \in I_k | X_j \leq G_j^{-1}(\alpha), \forall j \in \bar{I}_k)$ [13, 15]. Contrary to the case $d = 2$, this expression depends, in general, on the composition of I_k and not only on its cardinal. We will limit the study to a particular case of exchangeable copulas, for which $\lambda_{L,I_k} = \lambda_{L,I'_k}$ if $|I_k| = |I'_k|$. This assumption is in particular verified for Archimedean copulas [15], and we subsequently write $\lambda_{L,k}$ instead of λ_{L,I_k}. Symmetrically, one can define upper tail dependence parameters. For instance, for bivariate variables, it is $\lambda_U = \lim_{\alpha \to 1} \mathbb{P}(X_1 > G_1^{-1}(\alpha) | X_2 > G_2^{-1}(\alpha))$, if the limit exists.

Proposition 2.7. *Let R be defined as in equation (2.1) for an exchangeable copula C such that $\delta : v \in [0,1] \mapsto C(v, ..., v)$ is invertible. If all the lower and upper tail dependence parameters exist and are noted $\lambda_{L,k}$ and $\lambda_{U,k}$, for $k \in \{1, ..., d-1\}$, then the asymptotic difference $R(\alpha)$ between the probabilities associated to the lower- and upper-orthant quantiles is such that the following is applicable:*

$$\lim_{\alpha \to 0} \frac{R(\alpha)}{\alpha} \leq \frac{1}{\lambda_{L,d-1}} \sum_{k=1}^{d-1} \binom{d}{k} (1 - \lambda_{L,k}) \tag{2.2}$$

with equality only if the lower tail dependence parameters are all equal to 1, and

$$\lim_{\alpha \to 1} \frac{R(\alpha)}{1 - \alpha - R(\alpha)} \leq \frac{1}{\lambda_{U,d-1}} \sum_{k=1}^{d-1} \binom{d}{k} (1 - \lambda_{U,k}), \tag{2.3}$$

with equality only if the upper tail dependence parameters are all equal to 1.

The proof of Proposition 2.7 is reported in the Appendix.

Proposition 2.7 gives an upper bound for the difference of probability associated to the lower-orthant and the upper-orthant approaches. Naturally, the level of probability associated to the corresponding Kendall quantile is between lower-orthant and upper-orthant measures. In particular, $r(\alpha) \leq R(\alpha)$. This provides an upper bound for $r(\alpha)$. When $d = 2$, inequalities in Proposition 2.7 are simplified and upper bounds in equations (2.2) and (2.3) are $2(\lambda_L^{-1} - 1)$ and $2(\lambda_U^{-1} - 1)$ respectively. In special cases, if the lower tail dependence is strong, λ_L is close to 1 and the upper bound in equation (2.2) is close to 0: the lower-orthant, upper-orthant and Kendall quantile are very close in the lower tail. On the contrary, when the lower tail dependence is weak, λ_L is close to 0 and the upper bound in equation (2.2) tends to infinity: the lower-orthant, upper-orthant and Kendall quantiles are very disparate in the lower tail.

In Proposition 3.1, we use the result of Proposition 2.7 in the particular framework of Archimedean copulas with regularly varying generators.

3. KENDALL'S MULTIVARIATE QUANTILE FOR AN ARCHIMEDEAN COPULA

In this section, we assume that the multivariate distribution of the random vector X of dimension d is provided by an Archimedean copula C of generator ϕ:

$$C : (u_1, ..., u_d) \in [0,1]^d \mapsto \phi^{-1}\left(\sum_{j=1}^{d} \phi(u_j)\right).$$

It is a wide class of copulas which includes the following copulas: independent, Gumbel, Clayton, Frank, Joe, and Ali-Mikhail-Haq, among others. Moreover, this framework leads to simple expressions for the Kendall function, so that it is an interesting illustration of our theory.[2]

3.1. Theoretical results

In this Archimedean framework, we make some assumptions regarding ϕ.

Assumption 3.1. The generator ϕ is such that:

- $\phi : (0,1] \to [0,\infty)$,
- $\phi(1) = 0$,
- ϕ is strict, that is $\lim_{t\to 0^+} \phi(t) = \infty$,
- $(-1)^i(\phi^{-1})^{(i)}(x) > 0$ for all $1 \leq i \leq d$ and all $x \geq 0$,[3]
- $\lim_{t\to 0^+} \phi(t)^i(\phi^{-1})^{(i)}(\phi(t)) = 0$ for all $1 \leq i \leq d-1$.

[2] It is known that Archimedean copulas can be difficult to use in high dimensions for the purpose of estimation. Nevertheless, the vine approach permits bypassing this problem. Vine copulas are indeed based on nested bivariate copulas instead of a sole high-dimension copula [10, 25, 26]. Statistical selection techniques may help to truncate the vine so as to reduce the dimension of the problem in a relevant way [5]. For non-Archimedean copulas, semi-parametric methods may be used to estimate the Kendall function [35].

[3] In particular, ϕ is strictly decreasing and convex.

In Assumption 3.1, the fact that the generator is strict is intended to avoid that the zero curve of the copula may have a non-zero probability. The other assumptions are required by equation (3.1), which derives the Kendall distribution function in the Archimedean case with the help of the generator ϕ:

$$K : t \in (0,1] \mapsto t + \sum_{i=1}^{d-1} \frac{(-\phi(t))^i}{i!} (\phi^{-1})^{(i)}(\phi(t)), \tag{3.1}$$

where $f^{(i)}$ denotes the i-th derivative of f [2, 22]. We now apply this formula in two examples.

Example 3.1. The Gumbel copula is an Archimedean copula of parameter $\theta \geq 1$, generated by the function

$$\phi : t \mapsto (-\log(t))^\theta.$$

When $\theta = 1$, the Gumbel copula is equal to the independent copula. The inverse generator is $\phi^{-1}(x) = \exp(-x^{1/\theta})$. According to equation (3.1), if we consider the bivariate case, the Kendall function is as follows:

$$K : t \mapsto t - \frac{t \log(t)}{\theta}.$$

In Figure 2, we demonstrate how the Kendall function behaves when θ changes: the greater θ, the closer the Kendall function and the identity. In particular, when θ tends to infinity, K converges toward the identity, so that the lower-orthant Kendall quantile and the lower-orthant quantile are equal in this limit case.

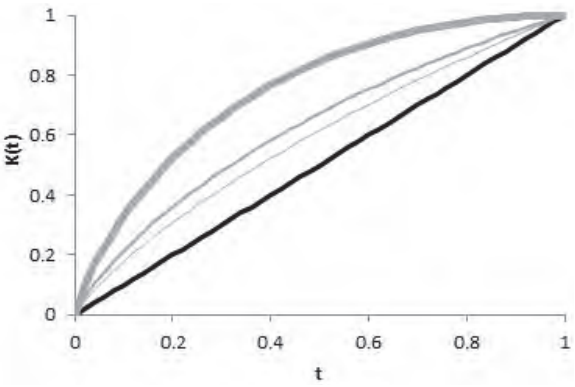


Figure 2: Kendall function (in grey) of the Gumbel copula for $\theta = 1$ (thick line), $\theta = 2$ (medium line), and $\theta = 3$ (thin line). The greater the difference between the Kendall function and the identity (in black, corresponding to $\theta \to \infty$), the greater the difference between the lower-orthant Kendall quantile and the lower-orthant quantile.

Example 3.2. The independent copula leads to easy formulas in higher dimensions. It is a particular case of the Gumbel copula with $\theta = 1$. According to equation (3.1), for a dimension $d \geq 2$, we get the following formula for K:

$$K : t \mapsto t\left(1 + \sum_{i=1}^{d-1} \frac{(-\log(t))^i}{i!}\right).$$

When d goes to infinity, $\sum_{i=1}^{d-1} \frac{(-\log(t))^i}{i!}$ tends toward $-1 + e^{-\log(t)} = -1 + 1/t$, for every $t \neq 0$.

Therefore, the limit behaviour of K, for $d \to \infty$, is a discontinuous function equal to 0 for $t = 0$ and equal to 1 everywhere else. It leads to the maximal difference possible between the lower-orthant Kendall quantile and the lower-orthant quantile. In Figure 3, we observe the Kendall function for various values of d.

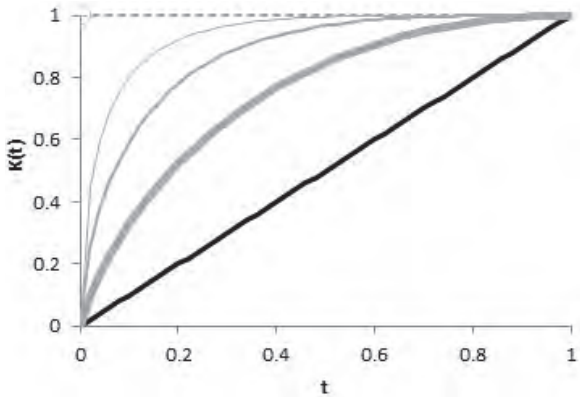


Figure 3: Kendall function (in grey) of the independent copula for $d = 2$ (thick line), $d = 3$ (medium line), $d = 4$ (thin line), and the limit case $d \to \infty$ (dotted line). The greater the difference between the Kendall function and the identity (in black), the greater the difference between the lower-orthant Kendall quantile and the lower-orthant quantile.

In Proposition 2.7, we saw that the probability distortion between the orthant and Kendall approaches was linked to the tail dependence. This is true, regardless of what the copula is. In the case of Archimedean copulas, we have an additional result allowing to link the probability distortion to the regular variations of the inverse generator. We first recall the definition of a regularly varying function:

Definition 3.1. A function f is regularly varying at 0, with index ρ, if

$$\forall s > 0, \lim_{x \to 0^+} \frac{f(sx)}{f(x)} = s^\rho.$$

Then, we note $f \in \mathcal{RV}_\rho(0)$.

Proposition 3.1. *If, for a given bivariate Archimedean copula, the inverse generator $\phi^{-1} \in \mathcal{RV}_{-\rho}(0)$, with $\rho > 0$, then the asymptotic difference $R(\alpha)$ between the probabilities associated to the lower- and upper-orthant quantiles, as defined in equation (2.1), is such that the following holds:*

$$\lim_{\alpha \to 0} \frac{R(\alpha)}{\alpha} \leq \sum_{k=1}^{d-1} \binom{d}{k} \left(d^{1/\rho} - (d-k)^{1/\rho} \right). \tag{3.2}$$

Proposition 3.1 is a direct consequence of Proposition 2.7 with Theorem 2.1 of lower tail coefficients in [15]. In equation (3.2), for bivariate variables, the upper bound is $2(2^{1/\rho} - 1)$. The faster ϕ^{-1} varies, the greater $|\rho|$ and the closer to zero is the upper bound of equation (3.2). On the contrary, slowly varying inverse generators are associated with a big probability distortion between lower-orthant, upper-orthant and Kendall approaches.

3.2. Simulation experiments

In this section, we apply the methodology presented above and evaluate the Kendall quantile using various Archimedean copulas. More precisely, we first illustrate the probability transformation implied by the Kendall distribution. Then, we present and compare orthant and Kendall quantiles with the help of simulations.

The different existing versions of Archimedean copulas are intended to depict various types of tail dependence [17]. Figure 4 shows how the Kendall distribution evolves with the type of copula we are using. It is interesting to note that the shape of the Kendall distribution obtained with a particular copula is consistent with the nature of the tail dependence of the copula. In other words, if the copula captures an upper-tail dependence behaviour, that is if extreme positive events have a tendency to occur simultaneously while others are independent, the Kendall distribution inflexion point is located in the left tail of the distribution. It is for instance the case of the Gumbel or, more sharply, of the Joe copula. If the copula captures a lower-tail dependence behaviour, as it is the case for the Clayton copula, the Kendall distribution inflexion point is located in the right tail of the distribution. The Frank copula is more body-centred, i.e. events present in the body are more dependent than those present in the tails. In this case, the twist in the Kendall function is similar for lower and upper tails.

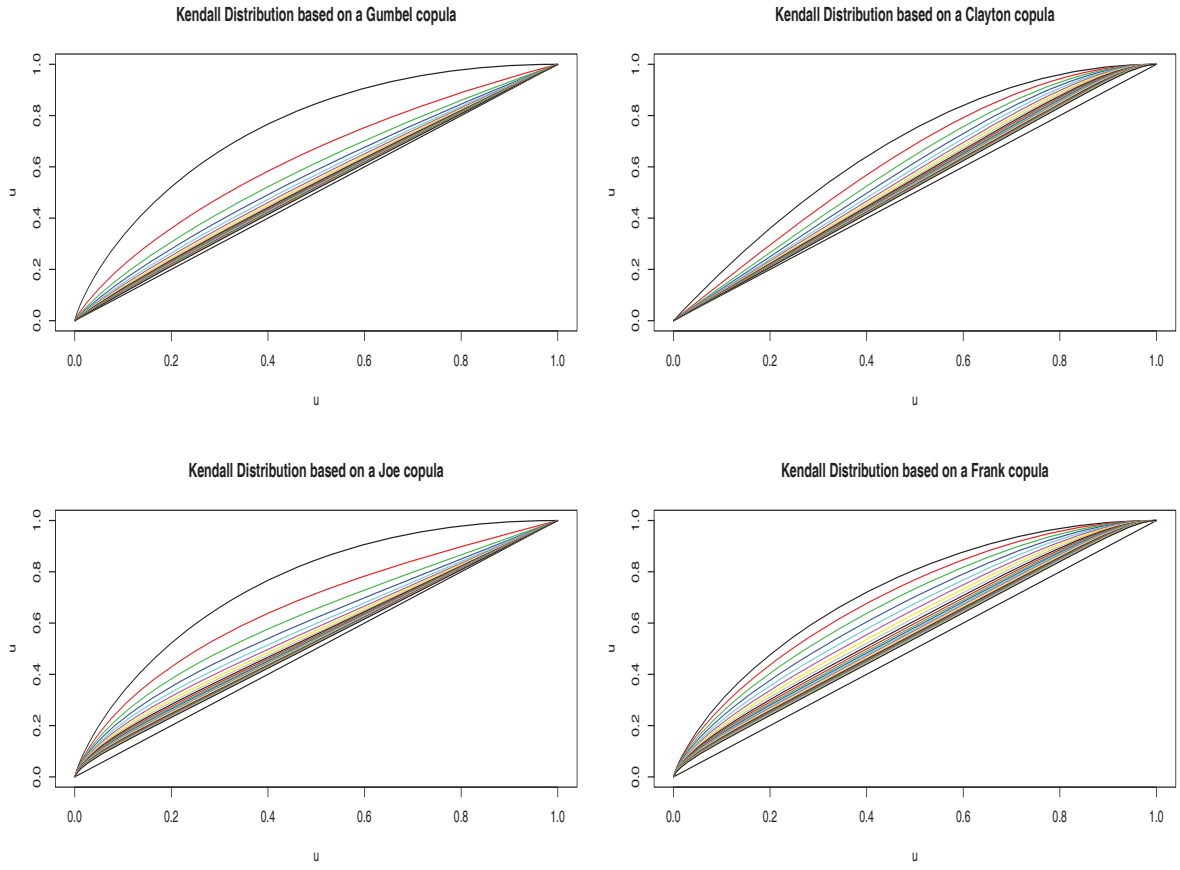


Figure 4: Kendall functions of four Archimedean copulas: Clayton, Gumbel, Frank, and Joe. The parameters of each copula are varying from 1 to 25. The bisector of the unit square corresponds to the orthant approach.

We now compare the quantiles obtained from the orthant and the Kendall approaches. Recall that multivariate quantiles will be represented by sets of vectors. To initiate our experimentation, we build a Clayton copula function, with parameter equal to 3. We used two lognormal marginal distributions, with the following sets of parameters: ($\mu = 5$, $\sigma = 2$) and ($\mu = 8$, $\sigma = 1.2$). Lower-orthant quantiles are obtained by calculating all the combinations of all pairs of margins providing the same bivariate probability. As analysed above, the Kendall distribution transforms the natural probabilities, taking into account the shape of the copula. This transformation allows us to calculate, in a similar fashion, the lower-orthant quantile and the Kendall quantile, transforming the lower-orthant percentile into the Kendall one.

Figure 5 shows lower-orthant, upper-orthant, and lower-orthant Kendall quantiles using the Clayton copula. In this figure, the dotted line, which is a set of vectors, represents the lower-orthant Kendall quantile, the continuous line located above[4] the dotted line represents the lower-orthant quantile, and the continuous line located below represents the upper-orthant quantile. In this figure, the quantile is given at the same percentile, but we see that the lower-orthant Kendall quantile is not equivalent to the lower-orthant quantile, as the Kendall function twists the probabilities. As a result, the lower-orthant quantile curve which is identical to the lower-orthant Kendall quantile curve of probability level α is below the lower-orthant quantile obtained at the same percentile α. The opposite is observed for upper-orthant quantiles. Indeed, Figure 5 shows the orthant quantiles obtained for α equal to 71%, as well as their Kendall equivalent, i.e. for $K(\alpha)$ also equal to 71%. To obtain this specific value for $K(\alpha)$, α has to be equal to 56%. In other results not displayed in figures, the lower-orthant Kendall quantile of probability level 86% or 98% is equal to the lower-orthant quantile of probability level 80% or 89%. This illustrates the probability distortion induced by the choice of copula and the Kendall function.

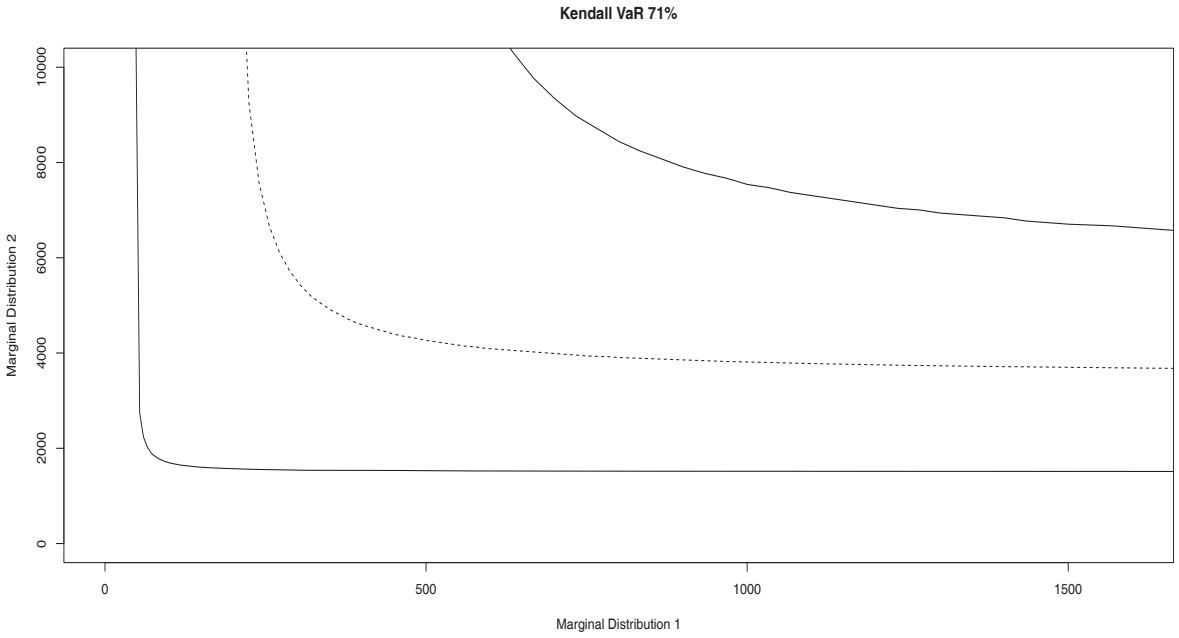


Figure 5: Lower-orthant Kendall quantile obtained using a Clayton copula with two different lognormal marginal distributions. The probability level is 71% for the Kendall quantile as well as for its lower-orthant (above) and upper-orthant (below) counterparts.

[4] The comparison between these sets of vectors must be understood in the sense of the lexicographical order for each pair of vectors.

4. CONCLUSION

In this paper, we focus on theoretical results on the concept of Kendall multivariate quantile, which allows for the calculation of a multivariate quantile using the probability transformation implied by the Kendall distribution. For instance, the Kendall distribution captures the intrinsic characteristics of the dependence architecture represented by the selected copula (non-linearity, upper or lower tail dependence etc.) and transfers it in one dimension. Therefore, the Kendall distribution allows the operation of a percentile transformation. We provide a simple relationship between the Kendall quantiles and the orthant quantiles, which allows to define the Kendall quantiles as a compromise between the bounds represented by both orthant quantiles. We also quantify the differences between the Kendall quantiles and the orthant quantiles, and link these asymptotic differences to tail dependence parameters.

For the orthant quantiles as well as for the Kendall quantiles, we observed that the non-linearity of the copulas implies that the sums of each set representing a given percentile are not constant. This phenomenon will have an important impact if any of these methodologies are used within financial institutions (for instance, banks or insurance companies), as if these approaches are used to evaluate the diversified capital pertaining to the various risks faced by them, the accurate value of the capital as well as the allocation of this capital will be problematic. Indeed, multiple sets of values will be representative of the same level of risks going from one end to the other.

In terms of applications, this result provides a variety of possible interpretations which will be the purpose of a companion paper, though as mentioned in introduction, it is an important topic considering the implications in terms of financial and climatic risks measurement.

A. APPENDIX – Proofs

Proof of Proposition 2.1: Let $L = \{y \in \mathbb{R}^d \mid K(G(y)) \leq \alpha\}$ and $U = \{y \in \mathbb{R}^d \mid \mathcal{K}(\overline{G}(y)) \leq 1-\alpha\}$. Let g be the probability measure associated with G. Then, by inversibility of the strictly monotonic K, $g(L) = \mathbb{P}[K(G(X)) \leq \alpha] = \mathbb{P}[G(X) \leq K^{-1}(\alpha)] = K(K^{-1}(\alpha)) = \alpha$ and $g(U) = 1-\alpha$. If $\overline{\Psi}^{\mathcal{K}}_{\alpha}(G) \cap \underline{\Psi}^{K}_{\alpha}(G) = \emptyset$, there are two possibilities: L and U are overlapping or they are not, but in both cases they have no boundary in common.

1. Either $L \cap U$ is an infinite and closed set. Since $\overline{\Psi}^{\mathcal{K}}_{\alpha}(G) \cap \underline{\Psi}^{K}_{\alpha}(G) = \emptyset$, every vector is at least in one of the two sets L and U. The probability measure of $L \cap U$ is then strictly positive, thanks to the assumptions regarding G, but it is contradictory to the fact that the measure of $\mathbb{R}^d$, equal to 1, is then $g(L) + g(U) - g(L \cap U) = \alpha + (1-\alpha) - g(L \cap U) = 1 - g(L \cap U)$.
2. Or the set S of vectors, defined by $S = \mathbb{R}^d \setminus (L \cup U)$, is an infinite and closed set. Then, similar to the previous case, $1 = g(L) + g(U) + g(S) = 1 + g(S)$, which is contradictory to the fact that its probability measure is expected to be strictly greater than 0. □

Proof of Proposition 2.2: Let $A = \{y \in \mathbb{R}^d \mid G(y) \geq \alpha\}$ and $B = \{y \in \mathbb{R}^d \mid K(G(y)) \geq K(\alpha)\}$.

- K is an increasing function, since it is a probability distribution function. Therefore, it follows immediately that $A \subset B$.
- The reciprocal inclusion does not hold in general. However, with the assumption of strict monotonicity of K in a neighbourhood V_α of α, the restriction of K to V_α is invertible. Let $y \in B$ and $\alpha' \in V_\alpha$ such that $\alpha' < \alpha$ (it does not exist if $\alpha = 0$ but this case is trivial). Therefore, $K(\alpha') < K(\alpha)$. Let's assume $y \notin A$. Then $G(y) < \alpha$. Two cases arise here. First, if $G(y) \leq \alpha'$, then $K(G(y)) \leq K(\alpha') < K(\alpha)$, which is contradictory to the assumption $y \in B$. Second, if $G(y) \in (\alpha', \alpha)$, then $G(y)$ is in V_α, so that $K(G(y)) < K(\alpha)$; the contradiction also holds. Therefore, the assumption $y \notin A$ was absurd, and we can conclude that $B \subset A$.
- Finally, $A = B$.

As a consequence, when considering the definition of both quantiles, we get the following:

$$\begin{aligned} \underline{\Psi}_{\alpha}(G) &= \partial\{y \in \mathbb{R}^d | G(y) \geq \alpha\} \\ &= \partial\{y \in \mathbb{R}^d | K(G(y)) \geq K(\alpha)\} \\ &= \underline{\Psi}^{K}_{K(\alpha)}(G). \end{aligned}$$

□

Proof of Proposition 2.4: We prove the second assertion, the proof for the first one being similar. Whatever y and z in $\underline{\Psi}^{K}_{\alpha''}(G)$, z cannot be in the interior of the upper orthant of y. Indeed, in such a case, $G(z) > G(y)$ or, if $G(z) = G(y)$, y would not be on the border of $A = \{x \in \mathbb{R}^d \mid K(G(x)) \geq \alpha''\}$, since all the lower orthant of z, in the interior of which is y, belongs to A.

Let $y \in \overline{\Psi}_{\alpha'}(G) \cap \underline{\Psi}^{K}_{\alpha''}(G)$. The probability measure of the upper orthant U of y is $1-\alpha'$. Since no vector of A is in the interior of U, the probability measure of A, which is equal to α'', is lower than the measure of the complement set of U, since G has no atoms. Therefore, $\alpha'' \leq 1 - (1-\alpha') = \alpha'$. □

Proof of Proposition 2.5: Kendall's tau and the Kendall function are linked by the following relation, for a continuous copula [22]:

$$\tau = \frac{2^d - 1 - 2^d \int_0^1 K(\alpha) d\alpha}{2^{d-1} - 1}.$$

Therefore:

$$\begin{aligned} \int_0^1 r(\alpha) d\alpha &= \int_0^1 (K(\alpha) - \alpha) d\alpha \\ &= \frac{2^d - 1 - (2^{d-1} - 1)\tau}{2^d} - \frac{1}{2} \\ &= (1 - \tau)\left(\frac{1}{2} - \frac{1}{2^d}\right). \end{aligned}$$

□

Proof of Proposition 2.6: By a change of variable, we have the following:

$$\begin{aligned} \int_0^1 \overline{r}(\alpha) d\alpha &= \int_0^1 (\mathcal{K}(\alpha) - \alpha) d\alpha \\ &= \int_0^1 \mathcal{K}(\alpha) d\alpha - \frac{1}{2}. \end{aligned}$$

Moreover, we note that $\mathcal{K}$ is, according to Definition 2.2, the Kendall function corresponding to the survival distribution function. It can thus be written in terms of the Kendall's tau of the survival copula, $\overline{\tau}$:

$$\int_0^1 \mathcal{K}(\alpha) d\alpha = \frac{2^d - 1 - (2^{d-1} - 1)\overline{\tau}}{2^d}.$$

Besides, we know that the Kendall's tau of the survival copula is equal to the Kendall's tau of the copula itself [23], so that $\overline{\tau} = \tau$. This immediately leads to the result stated in the proposition. □

Proof of Proposition 2.7: First, we look for $u \in [0, 1]$, such that $\alpha \underset{\alpha \to 0}{\sim} G(G_1^{-1}(u), ..., G_d^{-1}(u))$. By using the corresponding copula, this is equivalent to $\alpha \underset{\alpha \to 0}{\sim} C(u, .., u)$. By Bayes' rule, we thus should have

$$\alpha \underset{\alpha \to 0}{\sim} u\lambda_{L,d-1}.$$

Therefore, we define u as $\alpha / \lambda_{L,d-1}$.

Then, we define a vector $(x_1, ..., x_d) = (G_1^{-1}(u), ..., G_d^{-1}(u))$. Since $G(G_1^{-1}(u), ..., G_d^{-1}(u)) = \alpha$, this vector belongs to $\underline{\Psi}_\alpha(G)$. It also belongs to $\overline{\Psi}_{\alpha + R(\alpha)}(G)$, by the definition of $R(\alpha)$. Incidentally, the probability measure of the complement of the upper right quadrant of this vector $(x_1, ..., x_d)$, that is to say $\alpha + R(\alpha)$, is such that the following is applicable:

$$\begin{aligned} \alpha + R(\alpha) &= 1 - \mathbb{P}(X_1 > x_1, ..., X_d > x_d) \\ &= 1 - \mathbb{P}(G_1(X_1) > u, ..., G_d(X_d) > u) \\ &= 1 - \mathbb{P}(U_1 > u, ..., U_d > u), \end{aligned}$$

with $U_1, ..., U_d$ uniform variables linked by the same copula C as $X_1, ..., X_d$. Then

$$\begin{aligned}\alpha + R(\alpha) \;=\; & \mathbb{P}(U_1 \leq u, \ ..., \ U_d \leq u) \\ & + \sum_{i=1}^{d} \mathbb{P}(U_1 \leq u, \ ..., \ U_i > u, \ ..., \ U_d \leq u) \\ & + \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} \mathbb{P}(U_1 \leq u, \ ..., \ U_i > u, \ ..., \ U_j > u, \ ..., \ U_d \leq u) \\ & + \cdots\end{aligned}$$

whose asymptote, as $\alpha \to 0$, is $\sum_{k=0}^{d-1} \binom{d}{k} P_{d-k}$, where $P_k = \mathbb{P}(U_1 \leq u, \ ..., \ U_k \leq u, \ U_{k+1} > u, \ ..., \ U_d > u)$, owing to the assumption that the lower tail dependence parameter is constant for a given size of I_k, whatever the composition of the subset I_k.

Last, we observe that $P_d = C(u, ..., u) \underset{\alpha \to 0}{\sim} \alpha$ and that, for $k \geq 1$,

$$\begin{aligned}P_k \;&=\; \mathbb{P}(U_{k+1} > u, \ ..., \ U_d > u |\ U_1 \leq u, \ ..., \ U_k \leq u)\ \mathbb{P}(U_1 \leq u, \ ..., \ U_k \leq u) \\ &=\; \Big(1 - \mathbb{P}(U_{k+1} \leq u, \ ..., \ U_d \leq u |\ U_1 \leq u, \ ..., \ U_k \leq u)\Big)\ \mathbb{P}(U_1 \leq u, \ ..., \ U_k \leq u),\end{aligned}$$

according to Bayes' rule. The asymptote of P_k, as $\alpha \to 0$, is thus $(1 - \lambda_{L,d-k})\, \mathbb{P}(U_1 \leq u, ..., U_k \leq u) \leq (1 - \lambda_{L,d-k})u$ according to the upper Fréchet-Hoeffding bound, with equality only if $\lambda_{L,d-k} = 1$. The equality also holds if, when focusing on the Fréchet-Hoeffding inequality, the variables are comonotonic, which also implies that lower tail dependence parameters are equal to 1. This leads to equation (2.2).

Concerning equation (2.3), we observe that the upper tail dependence parameters of a random vector are equal to the lower tail dependence of the opposite of the vector. We can thus directly apply equation (2.2) to $(-X_1, ..., -X_d)$, for a probability level $\overline{\alpha}$, when it tends to 0, a difference function $\overline{R}$ and lower tail dependence parameters $\overline{\lambda}_{L,k} = \lambda_{U,k}$. For a vector $(-x_1, ..., -x_d)$ belonging to the lower-orthant quantile of level $\overline{\alpha}$ of the distribution of $(-X_1, ..., -X_d)$, if the probability measure of its upper orthant is α, then $\overline{\alpha} = 1 - (\alpha + R(\alpha))$, noting that this upper orthant is the lower orthant of $(x_1, ..., x_d)$. Moreover, $\overline{R}(\overline{\alpha})$ is 1 minus the probability measure of both the lower and upper orthants, therefore $\overline{R}(\overline{\alpha}) = 1 - \alpha - \overline{\alpha} = R(\alpha)$. □

REFERENCES

[1] Abdous, B. and Theodorescu, R. (1992). Note on the spatial quantile of a random vector, *Statistics and Probability Letters*, **13**(4), 333–336.

[2] Barbe, P.; Genest, C.; Ghoudi, K. and Rémillard, B. (1996). On Kendall's process, *Journal of Multivariate Analysis*, **58**(2), 197–229.

[3] Beirlant, J.; Caeiro, F. and Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes, *REVSTAT – Statistical Journal*, **10**(1), 1–31.

[4] Brechmann, E. (2014). Hierarchical Kendall copulas: properties and inference, *Canadian Journal of Statistics*, **42**(1), 78–108.

[5] Brechmann, E.; Czado, C. and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data, *Canadian Journal of Statistics*, **40**(1), 68–85.

[6] Chakraborty, B. (2001). On affine equivariant multivariate quantiles, *Annals of the Institute of Statistical Mathematics*, **53**(2), 380–403.

[7] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths, *Annals of Statistics*, **42**(3), 1203–1231.

[8] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data, *Journal of the American Statistical Association*, **91**(434), 862–872.

[9] Chebana, F. and Ouarda, T. (2011). Multivariate quantiles in hydrological frequency analysis, *Environmetrics*, **22**(1), 63–78.

[10] Cooke, R.M. (1997). *Markov and entropy properties of tree-and vine-dependent variables.* In: "Proceedings of the ASA section of Bayesian Statistical Science", 27.

[11] Cossette, H.; Mailhot, M.; Marceau, É. and Mesfioui, M. (2013). Bivariate lower and upper orthant value-at-risk, *European Actuarial Journal*, **3**(2), 321–357.

[12] Cousin, A. and Di Bernardino, E. (2013). On multivariate extensions of value-at-risk, *Journal of Multivariate Analysis*, **119**(1), 32–46.

[13] De Luca, G. and Rivieccio, G. (2012). *Multivariate tail dependence coefficients for Archimedean copulae.* In: "Advanced Statistical Methods for the Analysis of Large Data-Sets" (pp. 287–296), Springer, Berlin, Heidelberg.

[14] Dhar, S.S.; Chakraborty, B. and Chaudhuri, P. (2014). Comparison of multivariate distributions using quantile-quantile plots and related tests, *Bernoulli*, **20**(3), 1484–1506.

[15] Di Bernardino, E. and Rullière, D. (2016). On tail dependence coefficients of transformed multivariate Archimedean copulas, *Fuzzy Sets and Systems*, **284**, 89–112.

[16] Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks, *Journal of Multivariate Analysis*, **97**(2), 526–547.

[17] Flores, Y.S. (2016). General multivariate dependence using associated copulas, *REVSTAT – Statistical Journal*, **14**(1), 1–28.

[18] Fraiman, R. and Meloche, J. (1999). Multivariate L-estimation, *Test*, **8**(2), 255–317.

[19] Garcin, M. and Guégan, D. (2012). Extreme values of random or chaotic discretization steps and connected networks, *Applied Mathematical Sciences*, **6**(119), 5901–5926.

[20] Genest, C. and Favre, A. (2007). Everything you always wanted to know about copula modeling but were afraid to ask, *Journal of Hydrologic Engineering*, **12**(4), 347–368.

[21] Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, **88**(423), 1034–1043.

[22] Genest, C. and Rivest, L. (2001). On the multivariate probability integral transformation, *Statistics and Probability Letters*, **53**(4), 391–399.

[23] Georges, P.; Lamy, A.G.; Nicolas, E.; Quibel, G. and Roncalli, T. (2001). *Multivariate survival modelling: a unified approach with copulas*. Available at SSRN 1032559.

[24] de Gooijer, J.; Gannoun, A. and Zerom, D. (2006). A multivariate quantile predictor, *Communications in Statistics – Theory and Methods*, **35**(1), 133–147.

[25] Guégan, D. and Hassani, B.K. (2013). Multivariate VaRs for operational risk capital computation: a vine structure approach, *International Journal of Risk Assessment and Management*, **17**(2), 148–170.

[26] Guégan, D. and Maugis, P.-A. (2010). New prospects on vines, *Insurance Markets and Companies: Analyses and Actuarial Computations*, **1**(1), 4–11.

[27] Koltchinskii, V. (1997). *M*-estimation, convexity and quantiles, *The Annals of Statistics*, **25**(2), 435–477.

[28] Lalaurette, F. (2003). Early detection of abnormal weather conditions using a probabilistic extreme forecast index, *Quarterly Journal of the Royal Meteorological Society*, **129**(594), 3037–3057.

[29] Liu, R.; Parelius, J. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussions and rejoinder), *The Annals of Statistics*, **27**(3), 783–858.

[30] Mahalanobis, P. (1936). *On the generalized distance in statistics*. In: "Proceedings of the National Institute of Sciences (Calcutta)", **2**(1), 49–55.

[31] Nelsen, R.B. (2007). *An Introduction to Copulas*, Springer Science & Business Media.

[32] Nelsen, R.B.; Quesada-Molina, J.J.; Rodríguez-Lallena, J.A. and Úbeda-Flores, M. (2003). Kendall distribution functions, *Statistics and Probability Letters*, **65**(3), 263–268.

[33] Neuhäuser, D.; Hirsch, C.; Gloaguen, C. and Schmidt, V. (2015). Joint distributions for total lengths of shortest-path trees in telecommunication networks, *Annals of Telecommunications*, **70**(5-6), 221–232.

[34] Noyan, N. and Rudolf, G. (2013). Optimization with multivariate conditional value-at-risk constraints, *Operations Research*, **61**(4), 990–1013.

[35] Salvadori, G.; Durante, F. and Perrone, E. (2013). Semi-parametric approximation of Kendall's distribution function and multivariate return periods, *Journal de La Société Française de Statistique*, **154**(1), 151–173.

[36] Sato, M.; Ichiki, K. and Takeuchi, T. (2011). Copula cosmology: constructing a likelihood function, *Physical Review D*, **83**(2), 023501.

[37] Scherrer, R.; Berlind, A.; Mao, Q. and McBride, C. (2009). From finance to cosmology: the copula of large-scale structure, *The Astrophysical Journal Letters*, **708**(1), L9.

[38] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications, *Statistica Neerlandica*, **56**(2), 214–232.

[39] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation, *Journal of Nonparametric Statistics*, **22**(7), 915–936.

[40] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

[41] Thorarinsdottir, T.L.; Scheuerer, M. and Heinz, C. (2016). Assessing the calibration of high-dimensional ensemble forecasts using rank histograms, *Journal of Computational and Graphical Statistics*, **25**(1), 105–122.

[42] Torres, R.; Lillo, R.E. and Laniado, H. (2015). A directional multivariate value at risk, *Insurance: Mathematics and Economics*, **65**, 111–123.

[43] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function, *The Annals of Statistics*, **28**(2), 461–482.

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 97–114
https://doi.org/10.57805/revstat.v21i1.398

Integrating Jackknife into the Theil–Sen Estimator in Multiple Linear Regression Model

Authors: Tolga Zaman ✉
– Department of Statistics, Faculty of Science, Çankırı Karatekin University, Çankırı 18100, Turkey
tolgazaman@karatekin.edu.tr

Kamil Alakuş
– Department of Statistics, Faculty of Science and Arts, Ondokuz Mayıs University, Samsun 55139, Turkey

Received: November 2020 Revised: October 2021 Accepted: October 2021

Abstract:

- In this study, we provide Theil–Sen parameter estimators, which are in multiple linear regression model based on a spatial median, to be examined by the jackknife method. To obtain the proposed estimator, apply the jackknife to a multivariate Theil–Sen estimator (MTSE) from Dang *et al.* estimators, who proved that the MTSE estimator is asymptotically normal. Robustness, efficiency, and non-normality of the proposed estimator is tested with simulation studies. As a result, the proposed estimator is shown to be robust, consistent, and more efficient in multiple linear regression models with arbitrary error distributions. Also, it is seen that the proposed estimator reduces the effects of outliers even more and gives more reliable results. So, it is clearly observed that the proposed estimator improves the outcome of the multivariate Theil–Sen estimator. In addition, we support with the aid of numerical examples to these results.

Keywords:

- *jackknife; robustness; efficiency; Theil–Sen estimator; multiple linear regression; spatial median.*

AMS Subject Classification:

- 62F40, 62G05, 62G35, 62H12.

✉ Corresponding author.

1. INTRODUCTION

Regression analysis; including the cause-result relationship examines the relationship between dependent (response) and independent (predictor) variables. Parametric regression analysis is based on certain assumptions. The most important of these assumptions, the mathematical form of the relationship between the dependent and independent variable is known in advance. The least squares method is proposed to be the most useful for solving such problems. The estimator of slope by this method is referred as the least squares estimator (LSE), which is the best linear estimator under the means of minimum variance if the variance of the error term is finite. However, LSE is vulnerable to gross errors and is also inefficient for distributions with heavy tails. In this case, in order to make better estimations, regression methods which allow the linearity assumption in the parametric regression to be stretched are needed. These methods are non-parametric regression models known as regression models. Nonparametric regression analysis, is the method that is successful for some of the assumptions used in case of failure in order to provide valid parametric regression methods. Several non-parametric methods were explored in the last century, such as the Theil–Sen estimator (TSE) [41]; [37], and various M-estimators [21]; [18]; [42].

We consider a multiple linear regression model,

$$Y = \beta X + \epsilon \tag{1.1}$$

where (X,Y) is observable but ϵ is not. β is an unknown parameter and ϵ has an unknown cdf, F. The mean of ϵ may not be zero. X and ϵ are independent. Let $(x_1, y_1), ..., (x_n, y_n)$ be independent random observations from the above model. In the literature, some researchers assumed that x_i s are random variables [3] and others assumed that the distribution of Y_i is

$$F_i(y) = F(y - \beta x_i), \qquad i = 1, 2, ..., n$$

where $x_i s$ $i = 1, 2, ..., n$ are known non-identical constant [37].

Since β is a slope parameter, under the assumption that all x_i s are distinct, Theil [41] proposed an estimator of β defined as

$$\hat{\beta} = med\left[(s_{ij} \diagdown s_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad 1 \leq i \leq j = 1, ..., n)\right] \tag{1.2}$$

where med stands for median. The estimator was referred to as Theil's estimator in literature. Theil's estimator was not defined if there exist ties among x_i s, and was extended by Sen [37] as

$$\hat{\beta}_n = med\left[(s_{ij} \diagdown s_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad if x_i \neq x_j 1 \leq i \leq j = 1, ..., n)\right]. \tag{1.3}$$

The new estimator was referred to as TSE.

Consider a multiple linear regression with p $\geq$ 1:

$$Y_j = \beta_0 + X_j^T \beta + \epsilon_j; \quad j = 1, 2, ..., n. \tag{1.4}$$

Following the above procedure, first, $\varphi = (\beta_0, \beta^T)^T$ can be found as the solution of equation (1.4)

$$Y_j - \beta_0 - X_j^T \beta = 0, \quad \boldsymbol{l}_{k+1} = \{j_1, ..., j_{k+1}\} \tag{1.5}$$

where, $\boldsymbol{l}_{k+1}$ is a $(k+1)$ subsets of $\{1, ..., n\}$. That is to say, if matrix $(k+1) \times (k+1)$ matrix $(X_l : l \in \boldsymbol{l}_{k+1})$, is invertible. This estimation is described with $\hat{\varphi}_{l_{k+1}}$ to emphasize dependency to $k+1$ observations. Later, natural expanding of TSE from a simple linear regression model to a multivariate regression model becomes multivariate median as following:

$$\hat{\varphi}_n = Mmed\{\hat{\varphi}_{l_{k+1}} : \forall \boldsymbol{l}_{k+1}\} \tag{1.6}$$

where, it should be pointed out that $\hat{\varphi}_{l_{k+1}}$ is at the same time the least squares estimator of φ based on $k+1$ observations $\{(X_j, Y_j) : j \in \boldsymbol{l}_{k+1}\}$. In this perspective, t different arbitrary combination of $\{(X_j, Y_j) : j \in \boldsymbol{l}_t\}$ can be chosen which means, here, $k+1 \leq t \leq n$ and it construct least squares estimator of $\hat{\varphi}_{\boldsymbol{l}_t}$. Then, multivariate TSE $\hat{\varphi}_n$ of the parameter φ is naturally will be multivariate median of all possible least square estimators and is described as below:

$$\hat{\varphi}_n = Mmed\{\hat{\varphi}_{l_t} : \forall \boldsymbol{l}_t\}. \tag{1.7}$$

Least squares estimation is as follows:

$$\hat{\varphi}_l = (X_{\boldsymbol{l}}^T X_{\boldsymbol{l}})^{-1} X_{\boldsymbol{l}}^T Y_{\boldsymbol{l}}. \tag{1.8}$$

In multivariate Theil–Sen estimator (MTSE), regression coefficients through the application of combination as $\binom{n}{t}$ is estimated with least squares method [10]. After every combination, spatial median belonging of obtained regression coefficient is computed. Regression coefficients belonging of MTSE are estimated by calculating as the spatial median obtained.

TSE is, with 0.293 breakdown point, robust and has a limited effect function and high asymptotic efficiency. For this reason, it is competes well with other slope estimators [37, 11, 44]. When we explore the literature for asymptotic characteristics of TSE, Sen [37] examined the asymptotic normality of estimation when cumulative probability function is continuous and showed that it is super-efficient for discrete error term. Even though most of its good characteristics are interpreted clearly and a lot of statisticians tried to expand on it [29, 50]. Since TSE is formulated only for a simple linear model, it is underdeveloped and rarely used. While for TSE to be expanded to a multiple linear model is obvious and attractive, this is technically hard and is a case which slows down the generalization and exploration processes. Oja and Niinimaa [29] generalized Theil–Sen estimation, which is in simple linear regression, to multiple linear regression using Oja [29] median. Oja's median is a special case of spatial median. These studies about Theil–Sen estimation are important for future studies. Peng *et al.* [32] established the asymptotic distribution and robust consistency of Theil–Sen estimation when cumulative probability function of the error term comes arbitrarily from both continuous and discrete distribution. Asymptotic distribution and robust consistency of Theil–Sen estimation can be examined as follows. In literature, there are various studies about TSE. See, e.g., regression estimation with Theil–Sen regression under the measurement errors, Fernandes and Leblanc [15]; inverse regression estimation with the help

of Theil–Sen regression, Lavagnini *et al.* [23]; multivariate regional estimations using Theil–Sen estimator, Zhou and Serfling [50]; and asymptotic multiple linear regression estimation using Theil–Sen regression, Shen [38]. The Theil–Sen estimator has been widely acknowledged in several popular textbooks on nonparametric statistics and robust regression, see, e.g., Sprent [39], Hollander and Wolfe [19, 20], and Rousseeuw and Leroy [35], Wilcox [46]. It also has been extensively studied in the literature. Sen [37] and Wilcox [44] investigated its asymptotic relative efficiency to the least squares estimator. Akritas *et al.* [1] applied it to astronomy and Fernandes and Leblanc [15] to remote sensing. Wilcox [45] investigated some results on extensions and modifications of the Theil–Sen regression estimator. Wang [43] studied its asymptotic properties for model (1.1) with a random covariate. Wang [43] showed that TSE is strongly consistent, and obtain its asymptotic distribution, which may not be a normal distribution if F is not absolutely continuous. Many of its extensions can be found in the literature, for example, in censored data; for details, see, e.g., Akritas *et al.* [1], Jones [22], and Mount and Netanyahu [27]. Dang *et al.* [10] proposed the Theil–Sen estimators of parameters in a multiple linear regression model based on a multivariate median, generalizing the Theil–Sen estimator in a simple linear regression model. The sample mean of the bootstrap sample is known as the bagging estimator or smoothed bootstrap estimator. Empirically, bootstrapping with the bagging estimator often outperforms bootstrapping with the original estimator, especially when the asymptotic distribution is non-normal. See Breiman [5], Yang [48], and Efron [13]. See Büchlmann and Yu [6] and Friedman and Hall [16] for theory and references for the bagging estimator. Pelawa Watagoda and Olive [31] show that if $\sqrt{n}(T_n - \beta) \to N_p(0, \Sigma)$, then then under regularity conditions, $\sqrt{n}\left(\overline{T}^* - T_n\right) \to 0$, $\sqrt{n}(T_i^* - T_n) \to N_p(0, \Sigma)$ and $\sqrt{n}\left(\overline{T}^* - \beta\right) \to N_p(0, \Sigma)$. We are using a similar idea to bagging with the jackknife to produce the jackknife multivariate Theil–Sen estimator (JMTSE) estimator. In this paper, following Dang *et al.* [10], jackknife method which is one of the resampling methods is integrated in the multivariate Theil–Sen method (MTSE) and by doing this, a new estimator named jackknife multivariate Theil–Sen estimator (JMTSE) is offered. Robustness property of MTSE and is improved and became attractiveness for accomplishing well. In order to compare the proposed estimator with MTSE and LSE methods, various simulation studies are designed and results of multiple Theil–Sen estimation in multiple linear regression analysis are improved. Also, behaviors of these estimators are examined with two original data sets.

The remainder of the paper is organized as follows. In Section 2, we describe the properties of Theil–Sen estimator and spatial median. In Sections 3 and 4, we present jackknife method which is one of the resampling methods and we present some theoretical results of the jackknife method. In Section 5, we suggest a new estimator using jackknife method. In Section 6, we introduce the results for both simulations and real data set examples. In final section, we made conclusions about the obtained results.

2. STRONG CONSISTENCY AND ASYMPTOTIC NORMALITY PROPERTIES OF THEIL–SEN ESTIMATORS

In this section, it is stated the results on the strong consistency and the asymptotic distribution of TSE $\hat{\beta}_n(\varphi)$ under the assumption that X_is are random variables with $Var(X) > 0$.

Let ζ_0, $\left\{\varphi : \hat{\beta}_n(\varphi) = \beta\right\}$ (for all big n) be an event. That is to say, when $\varphi \in \zeta_0$ is $n > n_\varphi$ for each, $\hat{\beta}_n(\varphi) = \beta$ is so that there is a n_φ. The following theorem establishes an interesting property of the estimator

If F is continuous, then TSE $\hat{\beta}_n$is strongly consistent, that is $\hat{\beta}_n \to \beta$ [43].

In this section, we study the asymptotic distribution of the Theil–Sen estimator for both discontinuous and continuous error cdf F.

Firstly, F is assumed discontinuous. Then,

$$P\left(n^v\left(\hat{\beta}_n - \beta\right) \to 0\right) = 1, v \geq 0. \tag{2.1}$$

It has gotten the asymptotic behavior of $\hat{\beta}_n$ [43].

Now supposed that F is continuous. Denote the cdf of X_1 by G, the cdf of $X_1 - X_2$ by G_2 and the cdf of $\varepsilon_1 - \varepsilon_2$ by F_2. Then G_2 and F_2 are symmetric distribution function. Let

$$\mu(t) = \int [1 - 2F_2(xt)]dG_2(x) \tag{2.2}$$

and

$$\sigma^2 = \frac{1}{3}E\Big[(1 - G(X_1) - G(X_1))^2\Big]. \tag{2.3}$$

When G is continuous, $\sigma^2 = \frac{1}{9}$ ([43, 32]. Also, for further information about strong consistency and asymptotic distribution belonging of TSE estimation, please look into Wang [43]. Let's make statements about the spatial median.

Let W be a p-variate random vector with cdf F, $p > 1$. The spatial median (sm) of W minimizes the objective function:

$$D_F(w) = \frac{1}{n}\sum_{i=1}^{n}\{\|W_i - w\| - \|W_i\|\}; w \in R^d \tag{2.4}$$

where $\|.\|$ is the Euclidean form. Let $S(w) = w/\|w\|(w \neq 0)$ be the spatial sign function. The sample statistical spatial depth

$$D_F(w) = 1 - \left\|\frac{1}{n}\sum_{i=1}^{n} S(w - W_i)\right\|; w \in R^d. \tag{2.5}$$

The spatial median is the multivariate median defined by the spatial depth, which is any value that maximizes the sample depth,

$$\widehat{sm} = \arg\sup D_F(x); x \in R^d. \tag{2.6}$$

The estimate $\widehat{sm}$ is unique if the observations do not fall on a line. The spatial median has good efficiency properties. Möttönen *et al.* [28] for example calculated the asymptotic relative efficiencies. For the strong consistency and asymptotic normality of the spatial median (for see in detail information, Chaudhuri [9] and Bose [4].

3. JACKKNIFE METHOD

Jackknife method is defined as the method which minimizes sample error used to estimate population parameter. First definition of this method is made by Quenouille [33] and it is improved by Tukey using confidence interval approach [17]. Efron [12] contributed to the estimation of standard error and bias of the method. Martin and Roberts [25] proposed jackknife-after-bootstrap method which is developed in order to determine efficient observations. Jackknife method gives confidence intervals and decreases bias of estimation when known approaches are having hard time. Jackknife method is a resampling method in estimation of population parameters and developed in order to minimize sample error related to obtaining narrow confidence intervals. This method is considered as a statistical process that aims to reveal the relationship between variables in the data set in many fields that require parameter estimation [40]. Jackknife method does not consider the distribution belonging to variables in the data set and in this regard, it is known as a non-parametric statistical process. In parameter estimation process with this method, estimation is made by throwing out one observation in the sample each time. Thus, effect of deviated values is tried to be eliminate. The fundamental logic of the Jackknife method is to produce n different sample (sub-sample), each $(n-1)$ sized, by excluding each sample observation from the data set. The fundamental logic of the method bases on calculating sampling statistics from remaining observations through excluding an observation in data set. Thus, n different observations from n observations can be formed.

Let be $X = (x_1, x_2, ..., x_n)$ sample and $\hat{\theta} = s(X)$ our estimator. According to jackknife methods, when i. observation are excluded, new sample is defined as:

$$x_{(i)} = (x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n); i = 1, 2, ..., n. \tag{3.1}$$

Because of this, its estimator is also defined as:

$$\hat{\theta}_{(-i)} = s\big(x_{(i)}\big) \tag{3.2}$$

where $\hat{\theta}_{(.)}$ is the estimate of θ and calculated through the equation, $\hat{\theta}_{(.)} = \frac{\sum_{i=1}^{n} \hat{\theta}_{(-i)}}{n}$. Jackknife estimate of standard error is calculated as

$$\widehat{se}_{Jackk} = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{\theta}_{(-i)} - \hat{\theta}_{(.)}\right)^2}{n(n-1)}}. \tag{3.3}$$

Applying jackknife method in non-parametric regression can be thought as same logic. By excluding one observation from the current dependent and independent variables, non-parametric regression method is applied in current data. This process is repeated times of sample size.

The parameters of non-parametric regression methods using jackknife method are found as follows:

- Firstly, $(n-1)$ sized n different subsamples are formed by removing the observations from the data one by one.
- Regression coefficients belonging to non-parametric regression methods in interest are estimated for each formed subsample. This regression coefficients are called deleted slope coefficients and is indicated by $\beta_{i(-j)}, j = 1, 2, ..., n; i = 1, 2, ..., p$.
- Lastly, in order to obtain Jackknife estimator of intercept parameter value, mean of values $(y_j - \left(\hat{\beta}_{1(-j)}\right)x_1 + \left(\hat{\beta}_{2(-j)}\right)x_2 + ... + \left(\hat{\beta}_{p(-j)}\right)x_p)$ is estimated as estimate $\hat{\beta}_{0(-j)}$ as below:

$$\hat{\beta}_{0(-j)} = \frac{\sum_{j=1}^{n}\left(y_j - \left(\hat{\beta}_{1(-j)}\right)x_1 + \left(\hat{\beta}_{2(-j)}\right)x_2 + ... + \left(\hat{\beta}_{p(-j)}\right)x_p\right)}{n-1}. \tag{3.4}$$

- Mean value of these coefficients that obtained for each subsample are Jackknife estimators and expressed as below:

$$\hat{\beta}_i^{J*} = \frac{\sum_{j=1}^{n}\left(\hat{\beta}_{i(-j}\right)}{n}; i = 1, 2, ..., p; j = 1, 2, ..., n; \hat{\beta}_0^{J*} = \frac{\sum_{j=1}^{n}\hat{\beta}_{0(-j)}}{n}. \tag{3.5}$$

4. SOME PROPERTIES OF JACKKNIFE ESTIMATORS

In this section, the theoretical results of the jackknife estimation on unbiased, consistency, and asymptotic distribution are indicated.

For $j = 1, ..., n$ supposed that the point jackknife estimations of the parameters are $\hat{\boldsymbol{\beta}}^{J*}$ and determine the sampling distribution of these estimators. Here $\hat{\boldsymbol{\beta}}^{J*}$ is defined as in the following equation. Let us consider the $px1$ dimensional $\hat{\boldsymbol{\beta}}^{J*}$ jackknife estimator vector of $\boldsymbol{\beta}$ parameters:

$$\hat{\boldsymbol{\beta}}^{J*} = \frac{1}{n}\sum_{j=1}^{n}\hat{\boldsymbol{\beta}}_{(-j)}. \tag{4.1}$$

First, let's find the expected value of the estimator $\hat{\boldsymbol{\beta}}^{J*}$ in equation (4.1):

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}\sum_{j=1}^{n}E\left(\hat{\boldsymbol{\beta}}_{(-j)}\right), \tag{4.2}$$

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}E\left[\hat{\boldsymbol{\beta}}_{(-1)} + \hat{\boldsymbol{\beta}}_{(-2)} + \cdots + \hat{\boldsymbol{\beta}}_{(-n)}\right], \tag{4.3}$$

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}\left[E\left(\hat{\boldsymbol{\beta}}_{(-1)}\right) + E\left(\hat{\boldsymbol{\beta}}_{(-2)}\right) + \cdots + E\left(\hat{\boldsymbol{\beta}}_{(-n)}\right)\right], \tag{4.4}$$

$$E\left(\hat{\boldsymbol{\beta}}^{J*}\right) = \frac{1}{n}[\boldsymbol{\beta} + \boldsymbol{\beta} + \cdots + \boldsymbol{\beta}] = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{\beta} = \frac{n\boldsymbol{\beta}}{n} = \boldsymbol{\beta}. \tag{4.5}$$

The result clearly shows that the estimate of $\hat{\boldsymbol{\beta}}^{J*}$ is unbiased for the parameter vector of $\boldsymbol{\beta}$ [7, 2].

Then, let's find the variance–covariance of the estimator $\hat{\boldsymbol{\beta}}^{J*}$ in equation (4.1). The jackknife variance–covariance estimator of $\hat{\boldsymbol{\beta}}^{J*}$ then can be written as follows:

$$
\begin{aligned}
(4.6)\qquad V(\hat{\boldsymbol{\beta}}^{J*}) &= S \\
&= \frac{1}{n}\sum_{j=1}^{n}(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*})'(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*}) \\
&= \frac{1}{n}\begin{bmatrix} \sum_{j=1}^{n}(\hat{\beta}_{1(-j)} - \hat{\beta}_1^{J*})^2 & \cdots & \sum_{j=1}^{n}(\hat{\beta}_{1(-j)} - \hat{\beta}_1^{J*})(\hat{\beta}_{p(-j)} - \hat{\beta}_p^{J*}) \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n}(\hat{\beta}_{1(-j)} - \hat{\beta}_1^{J*})(\hat{\beta}_{p(-j)} - \hat{\boldsymbol{\beta}}_p^{J*}) & \cdots & \sum_{j=1}^{n}(\hat{\beta}_{p(-j)} - \hat{\beta}_p^{J*})^2 \end{bmatrix}_{(\boldsymbol{p}\times\boldsymbol{p})} \\
&= \begin{bmatrix} V(\hat{\beta}_1^{J*}) & \cdots & \mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right) & \cdots & V\left(\hat{\beta}_p^{J*}\right) \end{bmatrix}.
\end{aligned}
$$

If the expected value of both sides is taken in equation (4.6), equation (4.7) is obtained.

$$
\begin{aligned}
(4.7)\qquad E\left[V\left(\hat{\boldsymbol{\beta}}^{J*}\right)\right] &= E(S) \\
&= \begin{bmatrix} E\left[V\left(\hat{\beta}_1^{J*}\right)\right] & \cdots & E\left[\mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right)\right] \\ \vdots & \ddots & \vdots \\ E\left[\mathrm{Cov}\left(\hat{\beta}_1^{J*}, \hat{\beta}_p^{J*}\right)\right] & \cdots & E\left[V\left(\hat{\beta}_p^{J*}\right)\right] \end{bmatrix}_{(\boldsymbol{p}\times\boldsymbol{p})} \\
&= \begin{bmatrix} V(\beta_1) & \cdots & \mathrm{Cov}(\beta_1, \beta_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(\beta_1, \beta_p) & \cdots & V(\beta_p) \end{bmatrix} = \boldsymbol{\Sigma}.
\end{aligned}
$$

As a result, the sampling distribution of $\hat{\boldsymbol{\beta}}^{J*}$ in jackknife estimator is obtained asymptotically $\hat{\boldsymbol{\beta}}^{J*} \sim N(\boldsymbol{\beta}; \boldsymbol{\Sigma})$.

The equation (4.7) clearly indicates that the estimate of $V\left(\hat{\boldsymbol{\beta}}^{J*}\right)$ is unbiased asymptotically for $\boldsymbol{\Sigma}$.

Finally, let's examine the consistency of the jackknife estimator of $\hat{\boldsymbol{\beta}}^{J*}$. Firstly the estimator has the variance $V\left(\hat{\boldsymbol{\beta}}^{J*}\right)$ and then we can write it as follows:

$$
(4.8)\qquad \lim_{n\to\infty} V(\hat{\boldsymbol{\beta}}^{J*}) = \lim_{n\to\infty} S = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{j=1}^{n}\left(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*}\right)'\left(\hat{\boldsymbol{\beta}}_{(-j)} - \hat{\boldsymbol{\beta}}^{J*}\right)\right] \to 0.
$$

So, the equation (4.8) shows that the jackknife estimation $\hat{\boldsymbol{\beta}}^{J*}$ is consistent for the parameter vector $\boldsymbol{\beta}$. That is, $\hat{\boldsymbol{\beta}}^{J*} \to \boldsymbol{\beta}$ [36].

5. JACKKNIFE MULTIVARIATE THEIL–SEN ESTIMATOR (JMTSE) IN MULTIPLE LINEAR REGRESSION MODEL

JMTSE is a modification of the Theil–Sen estimator based on the jackknife method in multiple linear regression, a technique that narrows the confidence interval and reduces the effect of extreme values. In the analysis of proposed JMTSE method, subsamples, which are obtained through excluding each observation value sequentially from sample, are used instead of sample data. The algorithm steps for proposed JMTSE method is given below:

Step 1: Firstly, n different subsample, each $(n-1)$ sized, are formed by excluding each of n observation sequentially.

Step 2: Arbitrary number t is determined as $k+1 \leq t < n$. Here p indicates number of independent variable and n indicates sample size.

Step 3: Regression coefficients are calculated by applying LSE method to each possible $\binom{n-1}{t}$ combinations according to arbitrarily determined t value in order to estimate parameter estimation values, β_i, $(i = 1, 2, ..., p)$. If we express obtained regression coefficients as L_{ij}, regression parameter estimations are multivariate median of L_{ij} values. The multivariate median used here is spatial median. In other words, $\beta_{i(-j)} = Mmed(L_{ij})$ $(i = 1, 2, ..., p)$; $(j = 1, 2, ..., n)$. Since this process will be repeated for each subsample, n number of $\hat{\beta}_{i(-j)}$ will be calculated.

Step 4: To calculate $\hat{\beta}_{0(-j)}$ estimation, there are different alternatives. These are:

i) If error term has a symmetric distribution around zero, $\hat{\beta}_{0(-j)}$ estimation is calculated by calculating each possible $\left(y_i - \hat{\beta}_{i(-j)}x_i\right)$. In other words, $\hat{\beta}_{0(-j)} = med\left(y_i - \hat{\beta}_{i(-j)}x_i\right)$.

ii) $\hat{\beta}_{0(-j)}$ estimation can be calculated by averaging all possible $\left(y_i - \hat{\beta}_{i(-j)}x_i\right)$ values. That means $\hat{\beta}_{0(-j)} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{\beta}_{i(-j)}x_i\right)}{n-1}$.

iii) Alternatively, (β_0, β_i) values can be estimated simultaneously with multivariate median in a less restricted situation. In other words, it can be calculated using $(\beta_0, \beta_i) = smed(\beta_{0(-j)}, \beta_{i(-j)})$.

Step 5: Means of all these coefficients for each sub-sample are proposed JMTSE estimations. In other words, $\hat{\beta}_0 = \frac{\sum_{j=1}^{n}\hat{\beta}_{0(-j)}}{n}$ ve $\hat{\beta}_i = \frac{\sum_{j=1}^{n}\hat{\beta}_{i(-j)}}{n}$.

There are certain things to consider while estimating with MTSE and proposed JMTSE methods. For example, if independent variables are categorical, then subsamples selected according to arbitrary t might be zero valued. In this case, estimation values with the LSE method can take values like (L_{ij}) $\pm\infty$. So, uncertainty may occur in this analysis. If L_{ij} values, which is calculated based on arbitrary t, is excluded from analysis, there would be data loss and this decreases the reliability of analysis. There are suggestions about this problem in simple linear regression models in Erilli and Alakuş [14]. If these suggestions are applied to JMTSE method:

1. β_i is calculated by replacing regression coefficients in (L_{ij}) value, which is found infinite, with maximum or minimum values in each possible L_{ij} values (in other words, $\max(L_{ij})$ instead of $+\infty$ and $\min(L_{ij})$ instead of $-\infty$) which is calculated according to arbitrary m.
2. The median of the data can be placed instead of regression coefficients in L_{ij}, which is found infinite, after infinite values are excluded from the data.
3. The trimmed mean values can be placed instead of regression coefficients in (L_{ij}), which is found infinite.

Using these methods, regression parameter estimations are calculated and regression coefficients for proposed JMTSE estimation method can be calculated considering the algorithm above. There is not a certain conclusion about which method gives the best result. It is advised to the researcher that all estimation results belonging to the all possible methods need to be obtained and compared according to data structure in order to obtain the best result.

In this article, the results of estimation of regression coefficients belonging to JMTSE method $(\beta_0; \beta_i, (i = 1, 2, ..., p))$, which is proposed in the simulation and real data applications, are obtained and interpreted simultaneously. In other words, $\left(\hat{\beta}_0, \hat{\beta}_i\right) = smed\left(\hat{\beta}_{0(-j)}, \hat{\beta}_{i(-j)}\right)$ [49].

6. RESULTS AND DISCUSSION

We conducted some numerical examples with Monte Carlo simulations and two real dataset examples. We measured the robustness and efficiency of the coefficients in the simulations. In real dataset examples we checked the prediction accuracies of the models and sparseness of the regression coefficients. The datasets are compatible for our aims which consist of heavy-tailed errors. All implementations were performed in R software [34].

6.1. Calculation and simulation studies

We performed Monte Carlo simulations study to evaluate the efficiency of proposed method. To compare the performance of proposed estimator, we employed two techniques ordinary least squares (LSE), multivariate Theil–Sen estimator (MTSE) [10]. Simulation design was constructed similar to Dang *et al.* [10].

In this section, various simulation studies are made with regards to robustness and efficiency in order to examine the behavior of proposed method. Some samples are produced from multiple linear regression model $Y_i = 2.5 + 3X_{1i} + 1.5X_{2i} + \varepsilon_i$, where, $X_{1i} \sim N(0, 1)$, $X_{2i} \sim U(0, 1)$ and ε_i are produced from different distributions with different purposes.

In this study, with the help of sub-samples obtained by excluding each observation values from the sample, we take a random sample of size t from the whole sample between

$k+1 \leq t \leq n$ and calculate the LSE based this random sample. This process is repeated in such a way that it does not exceed the combination of $\binom{n}{t}$. Then, the spatial median of the obtained LSE estimators is calculated simultaneously. The mean of these coefficients calculated for each sub-sample are proposed JMTSE estimates. Breakdown point depends on the choice of t. The highest breakdown point is reached when t takes its minimal value $t = k+1$. Therefore $t = 3$ was taken in this study.

Firstly, let us examine the robustness of proposed method. Sample sizes $n = 20, 40, 80$ are produced from multiple linear regression model $Y_i = 2.5 + 3X_{1i} + 1.5X_{2i} + \varepsilon_i$ with distribution $\varepsilon_i \sim N(0,1)$. Obtained data set is polluted with outliers (X_i, Y_i) of regression model $Y_i = -10 - 20X_{1i} - 25X_{2i} + \varepsilon_i$. Here, n_1 and n_2 are, respectively, count of good ones and count of bad ones (outliers). When we examine Table 1 that without outliers, all the LSE, MTSE and JMTSE performed well. However, with the presence of outliers, the LSE's method has completely deteriorated and become useless. While MTSE's and JMTSE's performed well until the ratio of outliers reaches 35–40%. But, it is seen that JMTSE's gives results closer to the real parameter value. Also, in the polluted data, LSE method have distant values to real regression coefficients. As pollution rate increased, it is observed that regression coefficients obtained by JMTSE is better compared to regression coefficients obtained by MTSE. In other words, it is shown that regression coefficients obtained through JMTSE method, which is proposed by hybridizing the resampling method, jackknife method, with MTSE estimation that is in literature, are closer to the real regression coefficients. Therefore, as a result of this simulation study, it is seen that proposed JMTSE method gives sufficient contribution to the literature.

Table 1: Robustness.

	True Parameter (2.5, 3, 1.5)		
	LSE	MTSE	JMTSE
$n = 20$	(2.480, 3.001, 1.521)	(2.488, 3.043, 1.523)	(2.469, 3.001, 1.509)
$n = 30$	(2.487, 3.004, 1.530)	(2.466, 3.023, 1.494)	(2.488, 3.017, 1.501)
$n = 40$	(1.894, 2.242, 1.084)	(2.792, 2.544, 0.417)	(2.226, 2.202, 0.875)
$n_1 = 19$, $n_2 = 1$	(2.264, 1.862, −0.436)	(2.420, 2.864, 1.550)	(2.462, 2.961, 1.482)
$n_1 = 18$, $n_2 = 2$	(0.855, 0.949, −0.219)	(2.180, 3.144, 1.370)	(2.563, 3.001, 1.221)
$n_1 = 16$, $n_2 = 4$	(−0.732, −2.403, −3.249)	(1.210, 2.363, −0.849)	(0.826, 1.183, 1.178)
$n_1 = 14$, $n_2 = 6$	(−0.935, −3.601, −6.696)	(0.912, 0.480, −2.213)	(1.633, 1.667, −1.413)
$n_1 = 12$, $n_2 = 8$	(−0.348, −0.736, −0.860)	(1.644, −0.225, 0.369)	(−0.187, −0.188, −0.394)

Robust estimator may lose efficiency. To investigate the efficiency, a simulation is conducted as follows. The values of mean square error belonging to LSE's, MTSE's and JMTSE's value $\hat{\beta}$ with sample size $n = 10, 20, 25, 30, 35, 40, 45, 50$ distribution $\varepsilon_i \sim N(0,1)$, and various outlier value ratios is calculated using $MSE = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\left(\hat{\beta}_i - \beta^{true}\right)^2$, where $\kappa = 1000$, $\beta^{true} = (2.5, 3, 1.5)$ and $\hat{\beta}_i$ is the estimate for i-th sample.

In Table 2, LSE method gave worse results compared to other methods when all cases of outlier ratios and sample sizes are considered. When sample size is fixed and outlier ratios are increased, experimental mean squared errors belonging to examined methods increased. For example, when sample size is taken as 20 and outlier ratios are 5%, 10%, and 20% respectively, experimental mean squared errors are 4.8272 when 5%; 15.7596 when 10% and 26.1113 when

20% respectively. Similarly, experimental mean squared errors belonging to MTSE method are 0.1373 when 5%, 0.2152 when 10% and 0.7689 when 20%. Finally, mean squared errors belonging to JMTSE method are 0.1381 when 5%, 0.148 when 10% and 0.7158 when 20%. That is to say, when examined methods are evaluated, it is seen that experimental mean squared errors belonging to estimation methods increases when sample size is fixed and outlier ratios are increased. When regression coefficients belonging to examined methods and outlier ratios are fixed, mean values for all sample sizes are calculated. In n=10,20,25,...,50 sample size and 5% pollution rate, MSE values belonging to estimations are calculated as 7.7903 for LSE, 0.6270 for MTSE and 0.3512 for JMTSE when outlier ratio is 5%; 13.3294 for LSE, 3.1737 for MTSE and 1.6983 for JMTSE when 10%; and 31.2867 for LSE, 5.1085 for MTSE and 4.4401 for JMTSE when 20%. If it is looked carefully, the in equation, $MSE_{JMTSE} < MSE_{MTSE} < MSE_{LSE}$ exists between MSE's belonging to estimation methods which are examined when outlier ratios are fixed. In this regard, when pollution rate is fixed and sample size is n=10,20,25,...,50, it is seen that JMTSE estimation method gives more efficient results. In a same way, when sample size is fixed and outlier ratio is increased, again, JMTSE estimation method gives results closer to real regression coefficients compared to LSE and MTSE estimation methods. In other words, it is seen that JMTSE method is more efficient in estimating real parameters in these cases. In summary, it is seen that results belonging to Theil–Sen method are improved with the addition of jackknife method. It is also observed that when the sample size increases, the values decrease in the mean square error of the methods.

Table 2: Efficiency comparisons.

Sample Size	Outlier ratio %								
	5%			10%			20%		
	LSE	MTSE	JMTSE	LSE	MTSE	JMTSE	LSE	MTSE	JMTSE
10	28.80	0.4422	0.5483	32.0605	0.5326	0.5677	66.5179	4.2187	5.2909
20	4.8272	0.1373	0.1381	15.7596	0.2152	0.1484	26.1113	0.7689	0.7158
25	8.145	3.85	1.754	12.0568	0.1455	0.1468	31.2967	1.2035	0.8902
30	4.8467	0.0697	0.0698	10.3119	4.5238	5.8034	27.5454	0.3196	0.3271
35	4.2601	0.0499	0.0499	12.5515	9.9076	0.3733	25.7719	9.1562	9.4418
40	3.2691	0.0509	0.0509	8.734	4.5144	1.8785	26.3822	6.2388	2.8074
45	4.3719	0.0477	0.0478	7.9876	5.4876	4.6048	22.2376	8.6498	7.3089
50	3.8026	0.3682	0.1505	7.1731	0.0631	0.0632	24.4306	10.3123	8.7384
Mean	7.7903	0.6270	0.3512	13.3294	3.1737	1.6983	31.2867	5.1085	4.4401

Considering multiple linear regression model $Y_i = 2.5 + 3X_{1i} + 1.5X_{2i} + \varepsilon_i$, for sample sizes $n = 20, 30, 40$ and $n = 50$ generate 1000 samples with errors from $\varepsilon_i \sim N(0,1)$ $\varepsilon_i \sim t(u)$ with two different degrees of freedoms (df) $u = 1, 3$. The prediction accuracies are evaluated with mean square error (MSE) as the following:

$$MSE = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\left(\hat{\beta}_i - \beta^{true}\right)^2$$

where $\kappa = 1000$, $\beta^{true} = (2.5, 3, 1.5)$ and $\hat{\beta}_i$ is the estimate for i-th sample. As for relative efficiency (RE) of $\hat{\beta}$, it is obtained by dividing the MSE of the LSE by that of $\hat{\beta}$. In Tables 3 and 4, the values of MSE and RE are given.

The relative efficiencies and MSE values of the MTSE and JMTSE are computed with respect to LSE in Table 3. When we examine Table 3, under the Gaussian model, the finite sample RE values of MTSE and JMTS are about 12–58% and the 52–98% which are acceptable. However, it is found that RE values of the JMTSE is bigger than 1 for heavy tail distributions t with $df = 3$ and $df = 1$(Cauchy). Especially, under the Cauchy model, JMTSE is more efficient compared to LSE.

Table 3: MSE values and Relative efficiencies of the MTSE and JMTSE with respect to LSE for some continuous distributions.

		Normal			T3			T1 (Cauchy)		
		LSE	MTSE	JMTSE	LSE	MTSE	JMTSE	LSE	MTSE	JMTSE
$n = 20$	MSE	0.072785	0.126266	0.10913	0.941811	0.533084	0.534346	802.2187	2.513192	2.364063
	RE	1	0.576442	0.666953	1	1.76672	1.76255	1	319.2031	339.339
$n = 30$	MSE	0.047412	0.081497	0.069998	0.523894	0.372488	0.365735	169.8011	0.856768	0.802462
	RE	1	0.581769	0.677338	1	1.406473	1.432442	1	198.188	211.6002
$n = 40$	MSE	0.035762	0.097724	0.068763	0.451233	0.52053	0.351836	163.7145	4.880104	3.658022
	RE	1	0.365944	0.520069	1	0.866873	1.28251	1	33.54733	44.75491
$n = 50$	MSE	0.019061	0.156786	0.019393	0.319796	0.267127	0.254355	2.67E+07	3.51E−01	3.28E−01
	RE	1	0.121571	0.982851	1	1.197171	1.257285	1	76068376	81402439
Mean			0.411432	0.711803		1.309309	1.433697		19017232	20350759

The relative efficiencies and MSE values of the JMTSE are computed with respect to MTSE in Table 4. From the Table 4, it can be concluded that when the error comes from the heavily tail distributions t with $df = 3$, the JMTSE competes the MTSE, especially for Cauchy. That means, JMTSE is more efficient compared to MTSE.

Table 4: MSE values and Relative efficiencies of the JMTSE with respect to MTSE for some continuous distributions.

		T3		Cauchy	
		MTSE	JMTSE	MTSE	JMTSE
$n = 20$	MSE	0.533084	0.534346	2.513192	2.364063
	RE	1	0.99764	1	1.063082
$n = 30$	MSE	0.372488	0.365735	0.856768	0.802462
	RE	1	1.018464	1	1.067674
$n = 40$	MSE	0.52053	0.351836	4.880104	3.658022
	RE	1	1.479468	1	1.334083
$n = 50$	MSE	0.267127	0.254355	3.51E−01	3.28E−01
	RE	1	1.050214	1	1.070122
Mean			1.136446		1.13374

As a result, just like what simulation studies showed, it is found that proposed JMTSE method has more consistency than LSE and MTSE methods with regard to efficiency and robustness and results of MTSE method are improved.

6.2. Numerical illustrations

In this part we conducted some experiments on real datasets to evaluate the predictive performance of estimators. Meanwhile we presented the sparsity of the regression coefficients. For the application we used Coleman and Education expenditure data sets which are available in R software, "MASS" and "robustbase" packages [47, 24]. These datasets contain heavy-tailed errors so the real datasets are conformable for the computations. Coleman dataset contains 20 observations and 5 independent variables. This data set contains information on 20 schools from the Mid-Atlantic and New England states. The purpose is to predict the verbal mean test score [26]. Education expenditure data set consists of 50 observations and 3 independent variables. This data set is related with the education expenses of 50 states in the US. The aim is to predict the per capita expenditure for public education [8].

Predictive performance is measured by cross validation technique. The datasets are divided in two parts as test-train. Train sets contain 80% and test sets contain 20% of the datasets, respectively.

95% confidence intervals of the regression coefficients are given for LSE and JMTSE methods in Tables 5 and 6. From the Table 5, It is seen that independent variables, x_3, x_4 and x_5, are found significant for the model obtained through LSE method and all independent variables are found significant for the model obtained through proposed JMTSE method.

Table 5: Confidence intervals of the regression coefficients for the Coleman data set.

Coefficients	LSE		JMTSE	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Constant	16.743	59.624	35.175	37.619
x_1	−3.013	1.242	−0.675	−0.377
x_2	−0.004	0.159	0.058	0.086
x_3	0.540	0.835	0.654	0.684
x_4	0.323	1.619	0.881	0.975
x_5	−8.221	−1.629	−4.723	−4.342

In Table 6, independent variable x_2 is found significant for the model obtained through LSE method and independent variables, x_2 and x_3, are found significant for the model obtained through JMTSE method. As a result narrower confidence limits were estimated with the proposed estimator.

Table 6: Confidence intervals of the regression coefficients for the Education expenditure data set.

Coefficients	LSE		JMTSE	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Constant	−589.951	77.250	−346.571	−329.127
x_1	−0.113	0.134	−0.024	0.008
x_2	0.034	0.088	0.066	0.073
x_3	−0.142	1.625	0.879	0.954

The predictive performance of each approaches are given in Table 7. In experimental results, JMTSE performs better than LSE and MTSE in terms of prediction for new observations. It should be noted that the JMTSE proposed by using jackknife in MTSE is good for both data sets.

Table 7: Predictive performance results.

Methods	Coleman-MSE	Education E-MSE
LSE	16.423	4293.951
MTSE	15.209	3645.489
JMTSE	14.724	3564.373

7. CONCLUSION

Theil–Sen estimator is a point estimator of the slope parameter in the model and has many nice properties, including asymptotic normality. It has become a useful alternative solution for robust regression modelling with a high breakdown point and asymptotic efficiency. Although TSE has these many good properties, there are not many researches for Theil–Sen estimator in multiple linear regression methods. Jackknife method throwing an observation at a time from the sample which statistics calculates that as the number of individuals in the sample and the effect of extreme values can be defined as a method with relieving properties. The paper proposes a modification of the Theil–Sen estimator based on jackknife method. Simulation studies and real data applications is made in order to improve the results belonging to MTSE. Also, Jackknife estimations of parameters for Theil–Sen regression analysis, hypothesis test of parameters based on jackknife estimations and confidence intervals are examined.

According to simulations studies, robustness of methods investigated in the first phase of the simulation study is explored. When outliers do not exist in the data and distribution of error term is normal distribution, all of the LSE, MTSE and JMTSE estimation methods obtained results close to the real regression coefficients as a result of comparison with regard to robustness. Again, when the errors obtained from simulation study is normally distributed and the data is polluted with different ratios, LSE method got far away from real regression coefficients immediately and become unusable. However MTSE and JMTSE methods were able to endure until a certain point. It is seen that this ratio is around 35—40% for MTSE and JMTSE. Also JTSE is more capable of determining the real regression coefficients correctly. These findings exhibit the superiority of jackknife within MTSE in terms of variable selection. In the second phase of the simulation study, mean square error values are calculated for the methods which are investigated when errors are distributed normally but data has pollution with various ratios. As a result, it is clear that the proposed JMTSE method has a smaller mean square error than LSE and MTSE methods. In the third and fourth phase of the simulation study, no pollution is added to produced data. But when the error comes from the heavily tail distributions t with $df = 3$ and $df = 1$, it is found that JMTSE method, which is proposed with regards to sample sizes and arbitrary error term distributions, is more efficient than LSE and MTSE methods. Immediately after the simulation study, situations of methods

in interest are investigated considering two original datasets. According to estimation results, it is found that JMTSE method gives more efficient results than MTSE and LSE methods. As a result, robustness and effectiveness of MTSE is improved using jackknife method.

It is clearly seen that the proposed JMTSE method works well when n and t are small. This shows that the proposed JMTSE method is computationally feasible and the method is to have useful outlier resistance.

Consequently, a new estimator named JMTSE is proposed by integrating Jackknife method to Theil–Sen method in multiple linear regression. It is observed that proposed method reduces the effects of outliers even more and gives more reliable results. According to obtained results, resampling methods like Jackknife method can be applied in non-parametric regression methods successfully. Moreover, we demonstrated the applicability of jackknife with MTSE and concluded the success with several numerical examples. We suggest using JMTSE when there are many predictors for further practical studies to accomplish model selection in the presence of heavy-tailed errors.

ACKNOWLEDGMENTS

The authors would like to thank the TÜBİTAK (The Scientific and Technological Research Council of Turkey) which financially supporting this study.

REFERENCES

[1] Akritas, M.; Murphy, S. and LaValley, M. (1995). The Theil–Sen estimator with doubly censored data and applications to astronomy, *J. Amer. Statist. Assoc.*, **90**(429), 170–177.

[2] Bhushan, S. and Misra, P.K. (2017). On an unbiased jack-knife regression type estimator, *Advances in Fuzzy Mathematics*, **12**(1), 147–155.

[3] Bickel, P.J. (1982). On adaptive estimation, *Annals of Statistics*, **10**, 647–671.

[4] Bose, A. (1998). Bahadur representation of M_m estimates, *Ann. Stat.*, **26**, 771–777.

[5] Breiman, L. (1996). Bagging predictors, *Machine Learning*, **24**, 123–140.

[6] Büchlmann, P. and Yu, B. (2002). Analyzing bagging, *The Annals of Statistics*, **30**, 927–961.

[7] Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press.

[8] Chatterje, S. and Price, B. (1977). *Regression Analysis by Example*, John Wiley and Sons, New York.

[9] Chaudhuri, P. (1996). Multivariate location estimation using extension of R estimates through U-statistics type approach, *Ann. Statist.*, **20**, 897–916.

[10] Dang, X.; Peng, H.; Wang, X. and Zhang, H. (2008). *Theil–Sen Estimators in a Multiple Linear Regression Model*, Olemiss.edu, The University of Mississippi.

[11] Dietz, E.J. (1987). A comparison of robust estimators in simple linear regression, *Comm. Statist., Part B Simulation Comput.*, **16**, 1209–1227.

[12] Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans* (Vol. 38), SIAM.

[13] Efron, B. (2014). Estimation and accuracy after model selection, (with discussion), *Journal of the American Statistical Association*, **109**, 991–1007.

[14] Erilli, N.A. and Alakus, K. (2016). Parameter estimation in Theil–Sen regression analysis with jackknife method, *Eurasian Academy of Sciences Eurasian Econometrics, Statistics and Empirical Economics Journal*, **5**, 28–41.

[15] Fernandes R. and Leblanc, S.G. (2005). Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors, *Remote Sensing of Environment*, **95**(3), 303–316.

[16] Friedman, J.H. and Hall, P. (2007). On bagging and nonlinear estimation, *Journal of Statistical Planning and Inference*, **137**(3), 669–683.

[17] Gray, H.L. and Schucany, W.R. (1972). *The Generalized Jackknife Statistic* (Vol. 1), M. Dekker.

[18] Hampel, F.R. (1971). A general qualitative definition of robustness, *The Annals of Mathematical Statistics*, 1887–1896.

[19] Hollander, M. and Wolfe, D.A. (1973). *Nonparametric Statistical Methods*, first ed., Wiley, New York, 205–206.

[20] Hollander, M. and Wolfe, D.A. (1999). *Nonparametric Statistical Methods*, second ed., Wiley, New York, 421–423.

[21] Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, 799–821.

[22] Jones, M.P. (1997). A class of semi-parametric regressions for the accelerated failure time model, *Biometrika*, **84**, 73–84.

[23] Lavagnini, I.; Badocco, D.; Pastore, P. and Magno, F. (2011). Theil–Sen nonparametric regression technique on univariate calibration, *Inverse Regression and Detection Limits Talanta*, **87**, 180–188.

[24] Maechler, M.; Rousseeuw, P.; Croux, C.; Todorov, V.; Ruckstuhl, A.; Salibian-Barrera, M.; Verbeke, T.; Koller, M.; Conceição, E.L.T. and Palma, M.A. (2016). *robustbase: Basic Robust Statistics R package version* 0.92–7. http://CRAN.R-project.org/package=robustbase.

[25] Martin, M.A. and Roberts, S. (2010). Jackknife-after-bootstrap regression influence diagnostics, *Journal of Nonparametric Statistics*, **22**(2), 257–269.

[26] Mosteller, F. and Tukey, J.W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, MA.

[27] Mount, D.M. and Netanyahu, N.S. (2001). Efficient randomized algorithms for robust estimation of circular arcs and aligned ellipses, *Comput. Geometry: Theory Appl.*, **19**, 1–33.

[28] Möttönen, J.; Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests, *Annals of Statistics*, **25**, 542–552.

[29] Oja, H. (1983). Descriptive statistics for multivariate distributions, *Statist. Probab. Lett.*, **1**, 327–332.

[30] Oja, H. and Niinimaa, A. (1984). *On robust estimation of regression coefficients*, Research Report, Department of Applied Mathematics and Statistics, University of Oulu, Finland.

[31] Pelawa Watagoda, L.C.R. and Olive, D.J. (2019). *Bootstrapping Multiple Linear Regression After Variable Selection*, online at http://lagrange.math.siu.edu/Olive/ppboottest.pdf.

[32] Peng, H.; Wang, S. and Wang, X. (2008). Consistency and asymptotic distribution of the Theil–Sen estimator, *Journal of Statistical Planning and Inference*, **138**(6), 1836–1850.

[33] Quenouille, M.H. (1949). *Approximate tests of correlation in time-series 3*. In: "Mathematical Proceedings of the Cambridge Philosophical Society", **45**(3), 483–484, Cambridge University Press.

[34] R Core Team (2017). R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

[35] Rousseeuw, P.J. and Leroy, A.M. (2003). *Robust Regression and Outlier Detection*, Wiley, New York.

[36] Shao, J. and Wu, C.J. (1989). A general theory for jackknife variance estimation, *The Annals of Statistics*, 1176–1197.

[37] Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau, *J. Amer. Statist. Assoc.*, **63**, 1379–1389.

[38] Shen, G. (2009). Asymptotics of a Theil–Sen-type estimate in multiple linear regression, *Statistics & Probability Letters*, **79**(8), 1053–1064.

[39] Sprent, P. (1993). *Applied Nonparametric Statistical Methods*, second ed., CRC Press, New York.

[40] Swingler, D.N. (1995). On the application of the jackknife to the estimation of the parameters of short multisinusoidal signals using the data-matrix formulation, *IEEE Transactions on Signal Processing*, **43**(9), 2135–2143.

[41] Theil, H. (1950). *A rank invariant method of linear and polynomial regression analysis, III.* In: "Proc. of Koninklijke Nederl. Akad. Wetensch., Series A", **53**, 1397–1412.

[42] Tukey, J.W. (1977). *Exploratory Data Analysis*, MA, Addison-Wesley.

[43] Wang, X.Q. (2005). Asymptotics of the Theil–Sen estimator in simple linear regression models with a random covariate, *Nonparametric Statist.*, **17**, 107–120.

[44] Wilcox, R.R. (1998). Simulations on the Theil–Sen regression estimator with right-censored data, *Statist. Probab. Lett.*, **39**, 43–47.

[45] Wilcox, R.R. (2004). Some results on extensions and modifications of the Theil–Sen regression estimator, *British Journal of Mathematical and Statistical Psychology*, **57l**, 265–280.

[46] Wilcox, R.R. (2017). *Introduction to Robust Estimation and Hypothesis Testing*, 4th Edition, San Diego, CA, Academic Press.

[47] Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*, Fourth Edition, Springer, New York, ISBN 0-387-95457-0.

[48] Yang, Y. (2003). Regression with multiple candidate models: selecting or mixing?, *Statistica Sinica*, **13**, 783–809.

[49] Zaman, T. (2017). *Contributions to the jackknife estimation and test problems in multiple linear regression analysis*, Doctoral Thesis, Ondokuz Mayıs University, Graduate School of Sciences, 468099.

[50] Zhou, W. and Serfling, R. (2008). Multivariate spatial U-quantiles: a Bahadur–Kiefer representation, a Theil–Sen estimator for multiple regression, and a robust dispersion estimator, *Journal of Statistical Planning and Inference*, **138**(6), 1660–1678.

REVSTAT – Statistical Journal
Volume 21, Number 1, January 2023, 115–141
https://doi.org/10.57805/revstat.v21i1.408

Ordering Properties of the Smallest and Largest Order Statistics from Exponentiated Location-Scale Models Under Random Shocks

Authors: Molod Abdolahi
– Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Iran
molodabdolahi61@gmail.com

Gholamali Parham ✉
– Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Iran
parham_g@scu.ac.ir

Rahim Chinipardaz
– Department of Statistics, Faculty of Mathematical Sciences and Computer, Shahid Chamran University of Ahvaz, Ahvaz, Iran
chinipardaz_r@scu.ac.ir

Received: December 2020 Revised: November 2021 Accepted: November 2021

Abstract:

- In this paper, we discuss stochastic comparisons of lifetimes of series and parallel systems when the components are exponentiated location-scale models under random shocks. The results established here are developed in two directions. First, the comparisons are carried out with respect to usual stochastic ordering by using the concept of vector majorization for series and parallel systems. Next, when the matrix of parameters changes to another matrix of parameters in the sense of multivariate chain majorization, we study the usual stochastic order of the smallest order statistics when each component receives a random shock.

Keywords:

- *exponentiated location-scale family; matrix majorization; random shock; usual stochastic order; vector majorization.*

AMS Subject Classification:

- 60E15, 60K10.

✉ Corresponding author.

1. INTRODUCTION

In actuarial science, it is often of interest to compare stochastically extreme claim amounts from heterogeneous portfolios. In this regard, in the present work, we compare the extreme order statistics arising from two heterogeneous portfolios in the sense of the usual stochastic ordering. It is assumed that the portfolios belong to the general exponentiated location-scale model. The general exponentiated location-scale model includes several important statistical models such as generalized exponential distribution, generalized Weibull distribution, generalized Pareto distribution and many more. The exponentiated location-scale model has three types of parameters: location, scale and shape(or skewness) parameters. Location parameter is useful in modeling the insurance related data, since an insurance company suffers a claim from policyholder after a certain period of time from the date of beginning of the policy. Also, most of the data dealing with health care costs and economy are skewed. In finance, an investigator may often have small gains, but occasionally may have a few large losses. In this case, the data will invariably be negatively skewed. If, howeve, we have a reverse situation, the data in this case will be positively skewed. Thus, fitting skewed models to these types of data in finance and some other fields is a very important issue. in order to fit a skewed data, we require a model having a skewness parameter. The general exponentiated location-scale model possesses this kind of flexibility. For this reason, such a general model is useful for fitting various kinds of data sets. In practical situations, the extreme claim amounts play an important role as these provide useful information for determining annual premium. In actuarial science, it is an important issue in expressing preferences between random future gains or losses. In this direction, comparisons of claim amounts in two heterogeneous portfolio of risks based on different stochastic ordering such as usual become very useful. Order statistics have received a great amount of atention from various authors. It plays an important role in several areas of probability and statistics such as reliability theory, queueing theory, and survival analysis. Let $X_{1:n} \leq ... \leq X_{n:n}$ denote the order statistics corresponding to the random variables $X_1, ..., X_n$, where $X_{1:n}$ and $X_{n:n}$ corresponds to the sample minimum and sample maximum, respectively. The results of stochastic comparisons of the order statistics with independent and dependent sampling units can be seen in Dykstra *et al.* (1997) [9], Zhao and Balakrishnan (2011) [30], Li and Li (2015) [25], Torrado (2015) [28], Torrado and Kochar (2015) [29], Kochar and Torrado (2015) [18], Kundu *et al.* (2016) [24], Kundu and Chowdhury (2016 [19], 2018 [20], 2019 [21], 2020 [22]), Chowdhury and Kundu (2017 [4], 2018 [5]), Hazra *et al.* (2017) [14], Fang and Zhang (2013) [12], Fang and Xu (2019) [13], Das *et al.* (2020) [8], Kundu *et al.* (2022) [23], Chowdhury *et al.* (2022) [6], and the references there in for a variety of parametric models. The assumption in the papers lies in the fact that each of the order statistics $X_{1:n}, X_{2:n}, ..., X_{n:n}$ occurs with certainty and the comparison is carried out on the minimums or the maximums of the order statistics. Now, it may so happen that the order statistics experience random shocks which may or may not result in its occurrence and it is of interest to compare two such systems stochastically with respect to vector or matrix majorization. A random variable X is said to follow the exponentiated location-scale model if it's cumulative distribution function is given by

$$F_X(x; \lambda, \theta, \alpha) = \left[F\left(\frac{x-\lambda}{\theta}\right)\right]^{\alpha}, \qquad x > \lambda, \tag{1.1}$$

where $\lambda \in R$, $\alpha > 0$, $\theta > 0$ and F is the baseline distribution function. Here, λ, θ and α are respectively known as the location, scale, and shape parameters. We write $X \sim ELS(\lambda, \theta, \alpha)$

if X has the distribution function given by (1.1). The probability density function of the exponentiated location-scale model with (1.1) is denoted by f_X. In particular, when $\alpha = 1$, the model given in (1.1) reduces to the location-scale family of distributions. Further, when $\alpha = 1$ and $\lambda = 0$, (1.1) reduces to the scale family and when $\alpha = 1$ and $\theta = 1$, (1.1) becomes location family. The model in (1.1) coincides with the exponentiated-scale family when the location parameter λ is equal to 0. Thus (1.1) is a general family of distribution with great flexibility.

Let us assume series and parallel systems consist of n independent components in working conditions. Each component of the system receives a shock which may cause the component to fail. Let the random variable X_i denote lifetime of the i-th component in the system which experiences a random shock at binging. Also, suppose that I_{p_i} denotes independent Bernoulli random variables, independent of $X_i's$, with $E(I_{p_i}) = p_i$, where p_i will be called shock parameter hereafter. Then, the random shock impacts the i-th component ($I_{p_i} = 1$) with probability p_i or doesn't impact the i-th component ($I_{p_i} = 0$) with probability $1 - p_i$. Hence, the random variable $Y_i = I_{p_i} X_i$ corresponds to the lifetime of i-th component in a system under shock. Fang and Balakrishnan (2018) [10] have compared two such systems with generalized Birnbaum–Saunders components. Similar comparisons are carried out in the context of the insurance where largest or smallest claim amounts in a portfolio of risks are compared stochastically. One may refer to Barmalzan *et al.* (2017) [3], and Balakrishnan *et al.* (2018) [2].

The survival function of $Y_{1:n} = \min\{Y_1, ..., Y_n\}$ is given by

$$\bar{F}_{Y_{1:n}}(x; \underset{\sim}{p}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha}) = \prod_{i=1}^{n} p_i \left[1 - F^{\alpha_i}\left(\frac{x - \lambda_i}{\theta_i}\right)\right], \qquad x > \max\{\lambda_i,\, i = 1, ..., n\}, \tag{1.2}$$

and the cumulative distribution function of $Y_{n:n} = \max\{Y_1, ..., Y_n\}$ is given by

$$F_{Y_{n:n}}(x; \underset{\sim}{p}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha}) = \prod_{i=1}^{n} \left[1 - p_i \left[1 - F^{\alpha_i}\left(\frac{x - \lambda_i}{\theta_i}\right)\right]\right], \qquad x > \max\{\lambda_i,\, i = 1, ..., n\}, \tag{1.3}$$

where $\underset{\sim}{x} = (x_1, ..., x_n) \in I^n$ be any real vector and I^n denote a n-dimensional Euclidean space where $I \subseteq R$. Hereafter, we assume that $Y^*_{1:n}(Y^*_{n:n})$ denotes similarly the smallest (largest) order statistic arising from $Y^*_i = I_{p^*_i} X^*_i$, $i = 1, ..., n$, where $X^*_1, ..., X^*_n$ are independent non-negative random variables with $X^*_i \sim ELS(\lambda^*_i, \theta^*_i, \alpha^*_i)$, $i = 1, ..., n$ and $I_{p^*_1}, ..., I_{p^*_n}$ are independent Bernoulli random variables, independent of $X^*_i's$, with $E(I_{p^*_i}) = p^*_i$, $i = 1, ..., n$. Let

$$\begin{aligned}
P_n &= \left\{(\underset{\sim}{x}, \underset{\sim}{y}; n) :\, x_i > 0,\, y_j > 0 \text{ and } (x_i - x_j)(y_i - y_j) \leqslant 0,\; i, j = 1, ..., n\right\},\\
S_n &= \left\{(\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z}; n) :\, x_i, y_j, z_k > 0 \text{ and } x_i \leqslant (\geqslant) x_j,\, y_i \geqslant (\leqslant) y_j,\, z_i \geqslant (\leqslant) z_j\right\},\\
N_n &= \left\{(\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z}; n) :\, z_i \geq 1,\, x_i > 0,\, y_i > 0,\, x_i \leq (\geq) x_j,\, y_i \leq (\geq) y_j,\, z_i \leq (\geq) z_j\right\},\\
N^*_n &= \left\{(\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z}; n) :\, z_i > 0,\, x_i > 0,\, y_i > 0,\, x_i \leq (\geq) x_j,\, y_i \leq (\geq) y_j,\, z_i \leq (\geq) z_j\right\},\\
U_n &= \left\{(\underset{\sim}{w}, \underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z}; n) :\, w_i, x_j, y_k, z_l > 0,\, w_i \leqslant (\geqslant) w_j,\, x_i \geqslant (\leqslant) x_j,\, y_i \geqslant (\leqslant) y_j,\, z_i \geqslant (\leqslant) z_j\right\}.
\end{aligned}$$

The rest of this paper is organized as follows. In Section 2, we introduce some definitions and fundamental lemmas. In Section 3, we establish some ordering properties for the smallest and largest order statistics of the ELS model with associated random shocks. In Section 4, some special cases of our main results are added. Section 5 provides applications of the established results. Finally, in Section 6, we include some concluding.

2. PRELIMINARIES

In this section we provide some preliminary definitions and lemmas which will be useful in the sequel. To compare lifetimes of series and parallel systems, stochastic orders have been extensively used in the literature. Below, we present a few of them. Throughout the paper, we use the notations $R = (-\infty, +\infty)$, $R_+ = (0, +\infty)$ and $a \overset{\text{sign}}{=} b$ means that a and b have the same sign. Let X be non-negative random variable with distribution function F, and density function f. The survival function, hazard rate, and reversed hazard rate are $\bar{F} = 1 - F$, $r_X = \frac{f}{\bar{F}}$, and $\tilde{r}_X = \frac{f}{F}$, respectively.

Definition 2.1. Let X and Y be two absolutely continuous random variables with respective supports (l_X, u_X) and (l_Y, u_Y), where u_X and u_Y may be positive infinity, and l_X and l_Y may be negative infinity. Then, X is said to be smaller than Y in usual stochastic (st) order, denoted as $X \leqslant_{st} Y$, if $\bar{F}_X(t) \leqslant \bar{F}_Y(t)$ for all $t \in (-\infty, +\infty)$.

Let $\underset{\sim}{x} = (x_1, ..., x_n) \in I^n$ and $\underset{\sim}{y} = (y_1, ..., y_n) \in I^n$ be any two real vectors with $x_{(1)}, ..., x_{(n)}$ being the increasing arrangement of the components of the vector $\underset{\sim}{x}$. The following definitions may be found in Marshall *et al.* (2011) [27].

Definition 2.2.

(**i**) The vector $\underset{\sim}{x}$ is said to majorize the vector $\underset{\sim}{y}$ (written as $\underset{\sim}{x} \overset{m}{\geqslant} \underset{\sim}{y}$) if

$$\sum_{i=1}^{j} x_{(i)} \leqslant \sum_{i=1}^{j} y_{(i)}, \ j = 1, ..., n-1, \quad \text{and} \quad \sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)};$$

(**ii**) The vector $\underset{\sim}{x}$ is said to weakly supermajorize the vector $\underset{\sim}{y}$ (written as $\underset{\sim}{x} \overset{w}{\geqslant} \underset{\sim}{y}$) if

$$\sum_{i=1}^{j} x_{(i)} \leqslant \sum_{i=1}^{j} y_{(i)} \qquad \text{for} \quad j = 1, ..., n;$$

(**iii**) The vector $\underset{\sim}{x}$ is said to weakly submajorize the vector $\underset{\sim}{y}$ (written as $\underset{\sim}{x} \geqslant_w \underset{\sim}{y}$) if

$$\sum_{i=j}^{n} x_{(i)} \geqslant \sum_{i=j}^{n} y_{(i)} \qquad \text{for} \quad j = 1, ..., n;$$

(**iv**) The vector $\underset{\sim}{x}$ is said to be p-larger than the vector $\underset{\sim}{y}$ (written as $\underset{\sim}{x} \overset{p}{\geqslant} \underset{\sim}{y}$) if

$$\prod_{i=1}^{j} x_{(i)} \leqslant \prod_{i=1}^{j} y_{(i)} \qquad \text{for} \quad j = 1, ..., n;$$

(**v**) The vector $\underset{\sim}{x}$ is said to reciprocally majorize the vector $\underset{\sim}{y}$ (written as $\underset{\sim}{x} \overset{rm}{\geqslant} \underset{\sim}{y}$) if

$$\sum_{i=1}^{j} \frac{1}{x_{(i)}} \geqslant \sum_{i=1}^{j} \frac{1}{y_{(i)}} \qquad \text{for} \quad j = 1, ..., n.$$

It is not difficult to show that $\underset{\sim}{x} \overset{m}{\geqslant} \underset{\sim}{y} \Rightarrow \underset{\sim}{x} \overset{w}{\geqslant} \underset{\sim}{y} \Rightarrow \underset{\sim}{x} \overset{p}{\geqslant} \underset{\sim}{y} \Rightarrow \underset{\sim}{x} \overset{rm}{\geqslant} \underset{\sim}{y}$.

Definition 2.3. A function $\psi \colon I^n \to R$ is said to be schur-convex (schur-concave) on I^n if

$$\underset{\sim}{x} \overset{m}{\geqslant} \underset{\sim}{y} \quad \text{implies} \quad \psi(\underset{\sim}{x}) \geqslant (\leqslant) \psi(\underset{\sim}{y}) \quad \text{for all} \quad \underset{\sim}{x}, \underset{\sim}{y} \in I^n.$$

The following definitions related to matrix majorization may be found in Marshall *et al.* (2011) [27].

Definition 2.4.

(**i**) A square matrix Π_n, of order n, is said to be a permutation matrix if each row and column has a single entry 1, and all other entries as zero;

(**ii**) A square matrix $P = (p_{ij})$, of order n, is said to be doubly stochastic if $p_{ij} \geqslant 0$, for all $i, j = 1, ..., n$, $\sum_{i=1}^{n} p_{ij} = 1$, $j = 1, ..., n$ and $\sum_{j=1}^{n} p_{ij} = 1$, $i = 1, ..., n$;

(**iii**) A square matrix T_w, is said to be T-transform matrix if it has form $T_w = wI + (1 - w)\Pi$; where $0 \leqslant w \leqslant 1$, I is the identity matrix and Π is the permutation matrix. Let $T_{w_1} = w_1 I + (1 - w_1)\Pi_1$ and $T_{w_2} = w_2 I + (1 - w_2)\Pi_2$ be two transform matrices, where $0 \leq w_1, w_2 \leq 1$ and Π_1 and Π_2 are two permutation matrices that interchange two coordinates. Then, we say T_{w_1} and T_{w_2} have the same structure if $\Pi_1 = \Pi_2$, where Π_1 and Π_2 are permutation matrices with the same dimension, otherwise they are different structures.

Definition 2.5. Consider the $m \times n$ matrices $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ with rows $a_1, ..., a_m$ and $b_1, ..., b_n$, respectively.

(**i**) A is said to be larger than B in chain majorization, denoted by $A \gg B$, if there exists a finite set of $n \times n$ T-transform matrices $T_1, ..., T_k$ such that $B = AT_1 \cdots T_k$;

(**ii**) A is said to majorize B, denoted by $A > B$, if $A = BP$, where the $n \times n$ matrix P is doubly stochastic. Since a product of T-transforms is doubly stochastic, it follows that $A \gg B \Rightarrow A > B$;

(**iii**) A is said to be larger than the matrix B in row majorization, denoted by $A \overset{row}{>} B$, if $a_i \overset{m}{\geqslant} b_i$ for $i = 1, ..., m$. It is clear that $A > B \Rightarrow A \overset{row}{>} B$;

(**iv**) A is said to be larger than the matrix B in row weakly majorization, denoted by $A \overset{w}{>} B$, if $a_i \overset{w}{\geqslant} b_i$ for $i = 1, ..., m$. It is clear that $A \overset{row}{>} B \Rightarrow A \overset{w}{>} B$. Thus it can be written that $A \gg B \Rightarrow A > B \Rightarrow A \overset{row}{>} B \Rightarrow A \overset{w}{>} B$.

Also, we introduce the following notations.

(i) $D_+ = \{(x_1, ..., x_n) : x_1 \geqslant ... \geqslant x_n > 0\}$;

(ii) $\varepsilon_+ = \{(x_1, ..., x_n) : 0 < x_1 \leqslant ... \leqslant x_n\}$;

(iii) $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix}$.

The following two Lemmas are used to prove the two Theorems 2.17 and 2.18.

Lemma 2.1. *A differentiable function $\Phi : R_+^8 \to R_+$ satisfies*

$$\Phi(A) \geqslant (\leqslant)\Phi(B) \quad \text{for all } A, B \text{ such that } A \in U_2,\ A \gg B \tag{2.1}$$

if and only if

(i) $\Phi(A) = \Phi(A\Pi)$ *for all permutation matrice* Π *and for all* $A \in U_2$, *and*

(ii) $\sum_{i=1}^{4} (a_{ik} - a_{ij})[\Phi_{ik}(A) - \Phi_{ij}(A)] \geqslant (\leqslant)0$ *for all* $j, k = 1, 2$ *and for all* $A \in U_2$, *where* $\Phi_{ij}(A) = \frac{\partial \Phi(A)}{\partial a_{ij}}$.

Lemma 2.2. *Let the function $\gamma : R_+^4 \to R_+$ be differentiable and the functin $\Phi_n : R_+^{4n} \to R_+$ be defined as*

$$\Phi_n(A) = \prod_{i=1}^{n} \gamma(a_{1i}, a_{2i}, a_{3i}, a_{4i}).$$

Assume that Φ_2 satisfies (2.1). Then for $A \in U_n$ and $B = AT_w$, we have $\Phi_n(A) \geqslant (\leqslant)\Phi_n(B)$, where T_w is the T-transform matrix.

Proof: The proofs of Lemmas 2.1–2.2 are similar to those of Theorems 2 and 3 of Balakrishnan *et al.* (2015) [1], and Marshall and Olkin (1997) [26]. □

3. MAIN RESULTS

In this section we establish some ordering properties for the smallest and largest order statistics of the ELS model with associated random shocks. We now consider the following assumption.

Assumption 3.1. Suppose $X_1, ..., X_n$ are independent non-negative random variables with $X_i \sim ELS(\lambda_i, \theta_i, \alpha_i)$, and $I_{p_1}, ..., I_{p_n}$ are independent Bernoulli random variables, independent of $X_i's$, with $E(I_{p_i}) = p_i, i = 1, ..., n$. Further, suppose $X_1^*, ..., X_n^*$ are independent non-negative random variables with $X_i^* \sim ELS(\lambda_i^*, \theta_i^*, \alpha_i^*)$, and $I_{p_1^*}, ..., I_{p_n^*}$ are independent Bernoulli random variables, independent of $X_i^{*'}s$, with $E(I_{p_i^*}) = p_i^*, i = 1, ..., n$.

Theorem 3.1 shows that usual stochastic ordering holds between two parallel systems of heterogeneous components under random shocks for fixed θ and α.

Theorem 3.1. *Let Assumption 3.1 hold and $h : [0,1] \to R$ be a differentiable, increasing and strictly convex function. Also, $\lambda_i = \lambda_i^*$, $\theta_i = \theta_i^* = \theta$ and $\alpha_i = \alpha_i^* = \alpha$, where $i = 1, ..., n$. Then, $h(\underset{\sim}{p}) \geqslant_w h(\underset{\sim}{p}^*)$ implies $Y_{n:n} \geqslant_{st} Y_{n:n}^*$, provided $\underset{\sim}{\lambda} \in D_+$, $h(\underset{\sim}{p}) \in D_+$, and $h(\underset{\sim}{p}) = (h(p_1), ..., h(p_n))$.*

Proof: The cumulative distribution function of $Y_{n:n}$ is given by

$$F_{Y_{n:n}}(x) = \prod_{i=1}^{n} \left[1 - h^{-1}(u_i)\left[1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right]\right],$$

where $h(p_i) = u_i$. Let us define $\psi_1(\underset{\sim}{u}) = F_{Y_{n:n}}(x)$. Differentiating $\psi_1(\underset{\sim}{u})$, partially, with respect to u_i, we get

$$\frac{\partial \psi_1(\underset{\sim}{u})}{\partial u_i} = -\frac{\frac{dh^{-1}(u_i)}{du_i}\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)}{1 - h^{-1}(u_i)\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)}\psi_1(\underset{\sim}{u}) \leqslant 0, \tag{3.1}$$

so, $\psi_1(\underset{\sim}{u})$ is decreasing in each u_i. Again, it can be shown that

$$\frac{\partial \psi_1(\underset{\sim}{u})}{\partial u_i} - \frac{\partial \psi_1(\underset{\sim}{u})}{\partial u_j} \overset{\text{sign}}{=} \frac{\frac{dh^{-1}(u_j)}{du_j}\left(1 - F^{\alpha}\left(\frac{x-\lambda_j}{\theta}\right)\right)}{1 - h^{-1}(u_j)\left(1 - F^{\alpha}\left(\frac{x-\lambda_j}{\theta}\right)\right)} - \frac{\frac{dh^{-1}(u_i)}{du_i}\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)}{1 - h^{-1}(u_i)\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)} \tag{3.2}$$

Now,

$$\frac{\partial}{\partial u}\left(\frac{1}{1 - h^{-1}(u)(1 - F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}\right) = \frac{dh^{-1}(u)}{du}\left(1 - F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right) \geqslant 0,$$

implying that $\frac{1}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}$ is increasing in u. Thus, as $\underset{\sim}{\lambda} \in D_+$, $h(\underset{\sim}{p}) \in D_+$, for $i \leqslant j$ taking $\lambda_i \geqslant \lambda_j$ and $u_i \geqslant u_j$ and noticing that $h^{-1}(u)$ is increasing in u, it can be written that

$$\frac{1 - F^{\alpha}\left(\frac{x-\lambda_j}{\theta}\right)}{1 - h^{-1}(u_j)\left(1 - F^{\alpha}\left(\frac{x-\lambda_j}{\theta}\right)\right)} \leqslant \frac{1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)}{1 - h^{-1}(u_i)\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)},$$

Again, if $h(u)$ is convex in u, then $u_i \geqslant u_j$ gives $\frac{dh^{-1}(u_i)}{du_i} \geqslant \frac{dh^{-1}(u_j)}{du_j}$ which yields

$$\frac{\frac{dh^{-1}(u_j)}{du_j}\left(1 - F^{\alpha}\left(\frac{x-\lambda_j}{\theta}\right)\right)}{1 - h^{-1}(u_j)\left(1 - F^{\alpha}\left(\frac{x-\lambda_j}{\theta}\right)\right)} \leqslant \frac{\frac{dh^{-1}(u_i)}{du_i}\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)}{1 - h^{-1}(u_i)\left(1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta}\right)\right)}.$$

Substituting the above result in (3.2), we get $\frac{\partial \psi_1(\underset{\sim}{u})}{\partial u_i} - \frac{\partial \psi_1(\underset{\sim}{u})}{\partial u_j} \leqslant 0$. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], $\psi_1(\underset{\sim}{u})$ is Schur concave in $\underset{\sim}{u}$. Thus the result is proved by Theorem A.8 of Marshall *et al.* (2011) [27]. □

Theorem 3.2 shows that the majorized shape parameter vector leads to smaller systems lifetime in the sense of the usual stochastic ordering when the location and scale parameter vectors are constant and shock parameter vectors are heterogeneous.

Theorem 3.2. *Let Assumption 3.1 hold and $h : [0,1] \to R$ be a differentiable and decreasing function. Also, $\lambda_i = \lambda_i^* = \lambda$, $\theta_i = \theta_i^* = \theta$, and $\alpha_i^* = \beta_i$, where $i = 1, ..., n$. Then, $\underset{\sim}{\alpha} \overset{w}{\geqslant} \underset{\sim}{\beta}$ implies $Y_{n:n} \leqslant_{st} Y_{n:n}^*$, provided $\underset{\sim}{\alpha}, \underset{\sim}{\beta}, h(\underset{\sim}{p}) \in \varepsilon_+$.*

Proof: The cumulative distribution function of $Y_{n:n}$ is given by

$$F_{Y_{n:n}}(x) = \prod_{i=1}^{n} \left[1 - h^{-1}(u_i)\left[1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right]\right],$$

where $h(p_i) = u_i$. Differentiating $F_{Y_{n:n}}(x)$, partially, with respect to α_i, we get

$$\frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_i} = \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln[F\left(\frac{x-\lambda}{\theta}\right)]}{1 - h^{-1}(u_i)\left(1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right)} F_{Y_{n:n}}(x) \leqslant 0,$$

so, $F_{Y_{n:n}}(x)$ is decreasing in each α_i. Again, it can be shown that

(3.3)

$$\frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_i} - \frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_j} \overset{\text{sign}}{=} \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln[F\left(\frac{x-\lambda}{\theta}\right)]}{1 - h^{-1}(u_i)\left(1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right)} - \frac{h^{-1}(u_j)F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\ln[F\left(\frac{x-\lambda}{\theta}\right)]}{1 - h^{-1}(u_j)\left(1 - F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\right)}.$$

Now,

$$\frac{\partial}{\partial \alpha}\left(\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1 - h^{-1}(u)(1 - F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}\right) \overset{\text{sign}}{=} (1 - h^{-1}(u))F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right] \leqslant 0,$$

implying that $\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}$ is decreasing in α. Again, as $h(u)$ is decreasing in u, then

$$\frac{\partial}{\partial u}\left(\frac{h^{-1}(u)}{1 - h^{-1}(u)(1 - F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}\right) \overset{\text{sign}}{=} \frac{\partial h^{-1}(u)}{\partial u} \leqslant 0,$$

implying that $\frac{h^{-1}(u)}{1-h^{-1}(u)(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}$ is decreasing in u. Thus, as $\underset{\sim}{\alpha} \in \varepsilon_+$, $h(\underset{\sim}{p}) \in \varepsilon_+$, for $i \leqslant j$ taking $\alpha_i \leqslant \alpha_j$ and $u_i \leqslant u_j$ and noticing that $h^{-1}(u)$ is decreasing in u, it can be written that

$$\begin{aligned}\frac{h^{-1}(u_j)F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)}{1 - h^{-1}(u_j)\left(1 - F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\right)} &\leqslant \frac{h^{-1}(u_j)F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)}{1 - h^{-1}(u_j)\left(1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right)} \\ &\leqslant \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)}{1 - h^{-1}(u_i)\left(1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right)},\end{aligned}$$

which implies

$$\begin{aligned}\frac{h^{-1}(u_j)F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\ln[F\left(\frac{x-\lambda}{\theta}\right)]}{1 - h^{-1}(u_j)\left(1 - F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\right)} &\geqslant \frac{h^{-1}(u_j)F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln[F\left(\frac{x-\lambda}{\theta}\right)]}{1 - h^{-1}(u_j)\left(1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right)} \\ &\geqslant \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln[F\left(\frac{x-\lambda}{\theta}\right)]}{1 - h^{-1}(u_i)\left(1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right)}.\end{aligned}$$

Hence, substituting the above results in (3.3), we get $\frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_i} - \frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_j} \leqslant 0$. Thus, by Lemma 3.3 of Kundu *et al.* (2016) [24], it can be proved that $F_{Y_{n:n}}(x)$ is Schur-convex in $\underset{\sim}{\alpha}$. Thus the result is proved by Theorem A.8 of Marshall *et al.* (2011) [27]. □

Theorem 3.3. *Let Assumption 3.1 hold and $h : [0,1] \to R$ be a differentiable and decreasing function. Also, $\lambda_i = \lambda_i^* = \lambda$, $\theta_i = \theta_i^* = \theta$, and $\alpha_i^* = \beta_i$, where $i = 1, ..., n$. Then, $\underset{\sim}{\alpha}^{-1} \geqslant_w \underset{\sim}{\beta}^{-1}$ implies $Y_{n:n} \leqslant_{st} Y_{n:n}^*$, provided, $\underset{\sim}{\alpha}, \underset{\sim}{\beta}, h(\underset{\sim}{p}) \in \varepsilon_+$.*

Proof: The cumulative distribution function of $Y_{n:n}$ can be expressed as the function of a_i, where $a_i = \frac{1}{\alpha_i}$, $i = 1, ..., n$. We denote it by $\psi_2(\underset{\sim}{a})$, where $\underset{\sim}{a} = (a_1, ..., a_n)$, and

$$\psi_2(\underset{\sim}{a}) = \prod_{i=1}^{n} \left[1 - h^{-1}(u_i)\left[1 - F^{\frac{1}{a_i}}\left(\frac{x-\lambda}{\theta}\right)\right]\right],$$

where $h(p_i) = u_i$. It can be shown that the partial derivative of $\psi_2(\underset{\sim}{a})$, with respect to a_i increasing in $i = 1, ..., n$. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], it can be proved that $\psi_2(\underset{\sim}{a})$ is Schur-convex in $\underset{\sim}{a}$. Thus the result is proved by Theorem A.8 of Marshall *et al.* (2011) [27]. □

Theorem 3.4 shows that $Y_{n:n}$ is smaller than $Y_{n:n}^*$ with respect to the usual stochastic ordering when a vector of scale parameters is p-larger than that of another vector of the scale parameters with some additional conditions when the location and shape parameter vectors are constant and shock parameter vectors are heterogeneous. Similar result also hold under reciprocally majorized based conditions among the scale parameters.

Theorem 3.4. *Let Assumption 3.1 hold and $h : [0,1] \to R$ be a differentiable and decreasing function. Also, $\lambda_i = \lambda_i^* = \lambda$, $\theta_i^* = \delta_i$, and $\alpha_i = \alpha_i^* = \alpha$, where $i = 1, ..., n$. Then,*

(**i**) *$\underset{\sim}{\theta} \overset{p}{\geqslant} \underset{\sim}{\delta}$ implies $Y_{n:n} \leqslant_{st} Y_{n:n}^*$, provided $\underset{\sim}{\theta}, \underset{\sim}{\delta}, h(\underset{\sim}{p}) \in D_+$,, and $u\tilde{r}(u)$ is increasing in u;*

(**ii**) *$\underset{\sim}{\theta} \overset{rm}{\geqslant} \underset{\sim}{\delta}$ implies $Y_{n:n} \leqslant_{st} Y_{n:n}^*$, provided $\underset{\sim}{\theta}, \underset{\sim}{\delta}, h(\underset{\sim}{p}) \in \varepsilon_+$, and $\tilde{r}(u)$ is increasing in u.*

Proof:

(**i**): The cumulative distribution function of $Y_{n:n}$ can be expressed as the function of a_i, where $a_i = \ln \theta_i$, $i = 1, ..., n$. We denote it by $\psi_3(\underset{\sim}{a})$, where $\underset{\sim}{a} = (a_1, ..., a_n)$, and

$$\psi_3(\underset{\sim}{a}) = \prod_{i=1}^{n} \left[1 - h^{-1}(u_i)\Big(1 - F^{\alpha}(e^{-a_i}(x-\lambda))\Big)\right],$$

where $h(p_i) = u_i$. Differentiating $\psi_3(\underset{\sim}{a})$, partially, with respect to a_i, we get

$$\frac{\partial \psi_3(\underset{\sim}{a})}{\partial a_i} = -\frac{\alpha h^{-1}(u_i) e^{-a_i}(x-\lambda)\tilde{r}(e^{-a_i}(x-\lambda))F^{\alpha}(e^{-a_i}(x-\lambda))}{1 - h^{-1}(u_i)\left(1 - F^{\alpha}(e^{-a_i}(x-\lambda))\right)}\psi_3(\underset{\sim}{a}) \leqslant 0,$$

so, $\psi_3(\underset{\sim}{a})$ is decreasing in each a_i. Again, it can be shown that

$$\frac{\partial\psi_3(\underset{\sim}{a})}{\partial a_i}-\frac{\partial\psi_3(\underset{\sim}{a})}{\partial a_j}\overset{\text{sign}}{=}\frac{\alpha h^{-1}(u_j)e^{-a_j}(x-\lambda)\tilde{r}(e^{-a_j}(x-\lambda))F^{\alpha}(e^{-a_j}(x-\lambda))}{1-h^{-1}(u_j)\left(1-F^{\alpha}(e^{-a_j}(x-\lambda))\right)} - \frac{\alpha h^{-1}(u_i)e^{-a_i}(x-\lambda)\tilde{r}(e^{-a_i}(x-\lambda))F^{\alpha}(e^{-a_i}(x-\lambda))}{1-h^{-1}(u_i)\left(1-F^{\alpha}(e^{-a_i}(x-\lambda))\right)}. \tag{3.4}$$

Now,

$$\frac{\partial}{\partial a}\left(\frac{F^{\alpha}(e^{-a}(x-\lambda))}{1-h^{-1}(u)\left(1-F^{\alpha}(e^{-a}(x-\lambda))\right)}\right)\overset{\text{sign}}{=}-\alpha e^{-a}(x-\lambda)(1-h^{-1}(u))\times\tilde{r}(e^{-a}(x-\lambda)) \times F^{\alpha}(e^{-a}(x-\lambda))\leqslant 0,$$

implying that $\frac{F^{\alpha}(e^{-a}(x-\lambda))}{1-h^{-1}(u)\left(1-F^{\alpha}(e^{-a}(x-\lambda))\right)}$ is decreasing in a. Again, as $h(u)$ is decreasing in u, then Fang, L. and Balakrishnan, N. (2018) [10]. Ordering properties of the small25 est order statistics from generalized Birnbaum–Saunders models with associated 26 random shocks, Metrika, 81, 1, 19-35.

$$\frac{\partial}{\partial u}\left(\frac{h^{-1}(u)}{1-h^{-1}(u)\left(1-F^{a}(e^{-a}(x-\lambda))\right)}\right)\overset{\text{sign}}{=}\frac{\partial h^{-1}(u)}{\partial u}\leqslant 0,$$

implying that $\frac{h^{-1}(u)}{1-h^{-1}(u)\left(1-F^{a}(e^{-a}(x-\lambda))\right)}$ is decreasing in u. Thus, as $\underset{\sim}{\theta},h(\underset{\sim}{p})\in D_+$, for $i\leqslant j$ taking $a_i\geqslant a_j$ and $u_i\geqslant u_j$ and noticing that $h^{-1}(u)$ is decreasing in u, it can be written that

$$\frac{h^{-1}(u_j)F^{\alpha}(e^{-a_j}(x-\lambda))}{1-h^{-1}(u_j)\left(1-F^{\alpha}(e^{-a_j}(x-\lambda))\right)}\geqslant\frac{h^{-1}(u_j)F^{\alpha}(e^{-a_i}(x-\lambda))}{1-h^{-1}(u_j)\left(1-F^{\alpha}(e^{-a_i}(x-\lambda))\right)} \geqslant\frac{h^{-1}(u_i)F^{\alpha}(e^{-a_i}(x-\lambda))}{1-h^{-1}(u_i)\left(1-F^{\alpha}(e^{-a_i}(x-\lambda))\right)},$$

As $u\tilde{r}(u)$ is increasing in u, then

$$\frac{\alpha h^{-1}(u_j)e^{-a_j}(x-\lambda)\tilde{r}(e^{-a_j}(x-\lambda))F^{\alpha}(e^{-a_j}(x-\lambda))}{1-h^{-1}(u_j)\left(1-F^{\alpha}(e^{-a_j}(x-\lambda))\right)} \geqslant\frac{\alpha h^{-1}(u_i)e^{-\alpha_i}(x-\lambda)\tilde{r}(e^{-a_i}(x-\lambda))F^{\alpha}(e^{-a_i}(x-\lambda))}{1-h^{-1}(u_i)\left(1-F^{\alpha}(e^{-a_i}(x-\lambda))\right)}.$$

Hence, from (3.4), we get $\frac{\partial\psi_3(\underset{\sim}{a})}{\partial a_i}-\frac{\partial\psi_3(\underset{\sim}{a})}{\partial a_j}\geqslant 0$. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], it can be proved that $\psi_3(\underset{\sim}{a})$ is Schur-convex in $\underset{\sim}{a}$. Thus the result is proved by Lemma 3.1 of Khaledi *et al.* (2002) [16].

(**ii**): The cumulative distribution function of $Y_{n:n}$ can be expressed as the function of $b_i=\frac{1}{\theta_i}$, $i=1,...,n$. We denote it by $\psi_4(\underset{\sim}{b})$, where $\underset{\sim}{b}=(b_1,...,b_n)$:

$$\psi_4(\underset{\sim}{b})=\prod_{i=1}^{n}\left[1-h^{-1}(u_i)\left(1-F^{\alpha}(b_i(x-\lambda))\right)\right],$$

where $h(p_i)=u_i$. Differentiating $\psi_4(\underset{\sim}{b})$, partially, with respect to b_i, we see that $\psi_4(\underset{\sim}{b})$ is increasing in each $i=1,...,n$. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], it can be proved that $\psi_4(\underset{\sim}{b})$ is Schur-convex in $\underset{\sim}{b}$. Thus the result is proved by Lemma 4.1 of Hazra *et al.* (2017) [14]. □

Theorem 3.5 shows that the majorized shape parameter vector leads to smaller systems lifetime in the sense of the usual stochastic ordering when the location and scale and shock parameter vectors are heterogeneous.

Theorem 3.5. *Let Assumption 3.1 hold and $h:[0,1]\to R$ be a differentiable and decreasing function. Also, $\lambda_i=\lambda_i^*$, $\theta_i=\theta_i^*$, and $\alpha_i^*=\beta_i$, where $i=1,...,n$. Then, $\underset{\sim}{\alpha}\overset{w}{\geqslant}\underset{\sim}{\beta}$ implies $Y_{n:n}\leqslant_{st}Y_{n:n}^*$, provided $\underset{\sim}{\alpha},\underset{\sim}{\beta},h(\underset{\sim}{p}),\underset{\sim}{\lambda},\underset{\sim}{\theta},\in\varepsilon_+$.*

Proof: The cumulative distribution function of $Y_{n:n}$ is given by

$$F_{Y_{n:n}}(x)=\prod_{i=1}^{n}\left[1-h^{-1}(u_i)\left[1-F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\right]\right],$$

where $h(p_i)=u_i$. Differentiating $F_{Y_{n:n}}(x)$, partially, with respect to α_i, we get

$$\frac{\partial F_{Y_{n:n}}(x)}{\partial\alpha_i}=\frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\ln\left[F\left(\frac{x-\lambda_i}{\theta_i}\right)\right]}{1-h^{-1}(u_i)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\right)}F_{Y_{n:n}}(x)\leqslant 0,$$

so, $F_{Y_{n:n}}(x)$ is decreasing in each α_i. Again, it can be shown that

$$\begin{aligned}\frac{\partial F_{Y_{n:n}}(x)}{\partial\alpha_i}-\frac{\partial F_{Y_{n:n}}(x)}{\partial\alpha_j}\overset{\text{sign}}{=}&\frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\ln\left[F\left(\frac{x-\lambda_i}{\theta_i}\right)\right]}{1-h^{-1}(u_i)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\right)}\\&-\frac{h^{-1}(u_j)F^{\alpha_j}\left(\frac{x-\lambda_j}{\theta_j}\right)\ln\left[F\left(\frac{x-\lambda_j}{\theta_j}\right)\right]}{1-h^{-1}(u_j)\left(1-F^{\alpha_j}\left(\frac{x-\lambda_j}{\theta_j}\right)\right)}.\end{aligned}$$

Now,

$$\frac{\partial}{\partial\alpha}\left(\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}\right)\overset{\text{sign}}{=}(1-h^{-1}(u))F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]\leqslant 0,$$

implying that $\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}$ is decreasing in α. Again, as $h(u)$ is decreasing in u, then,

$$\frac{\partial}{\partial u}\left(\frac{h^{-1}(u)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}\right)\overset{\text{sign}}{=}\frac{\partial h^{-1}(u)}{\partial u}\leqslant 0,$$

implying that $\frac{h^{-1}(u)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}$ is decreasing in u. Again,

$$\frac{\partial}{\partial\lambda}\left(\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}\right)\overset{\text{sign}}{=}-(\frac{\alpha}{\theta})\tilde{r}\left(\frac{x-\lambda}{\theta}\right)F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\left(1-h^{-1}(u)\right)\leq 0,$$

implying that $\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}$ is decreasing in λ. Also,

$$\begin{aligned}\frac{\partial}{\partial\theta}\left(\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}\right)\overset{\text{sign}}{=}&-(\frac{\alpha}{\theta^2})(x-\lambda)\tilde{r}\left(\frac{x-\lambda}{\theta}\right)F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\\&\times\left(1-h^{-1}(u)\right)\leq 0,\end{aligned}$$

implying that $\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-h^{-1}(u)\left(1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\right)}$ is decreasing in θ. Thus, as $\underset{\sim}{\alpha}, \underset{\sim}{\beta}, h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \in \varepsilon_+$, for $i \leq j$ taking $\alpha_i \leq \alpha_j$, $u_i \leq u_j$, $\lambda_i \leq \lambda_j$, $\theta_i \leq \theta_j$ and noticing that $h^{-1}(u)$ is decreasing in u, it can be written that

$$\begin{aligned}
\frac{h^{-1}(u_j)F^{\alpha_j}\left(\frac{x-\lambda_j}{\theta_j}\right)}{1-h^{-1}(u_j)\left(1-F^{\alpha_j}\left(\frac{x-\lambda_j}{\theta_j}\right)\right)} &\leqslant \frac{h^{-1}(u_j)F^{\alpha_i}\left(\frac{x-\lambda_j}{\theta_j}\right)}{1-h^{-1}(u_j)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_j}{\theta_j}\right)\right)} \\
&\leqslant \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda_j}{\theta_j}\right)}{1-h^{-1}(u_i)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_j}{\theta_j}\right)\right)} \\
&\leqslant \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_j}\right)}{1-h^{-1}(u_i)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_j}\right)\right)} \\
&\leqslant \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)}{1-h^{-1}(u_i)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\right)},
\end{aligned}$$

which implies

$$\frac{h^{-1}(u_j)F^{\alpha_j}\left(\frac{x-\lambda_j}{\theta_j}\right)}{1-h^{-1}(u_j)\left(1-F^{\alpha_j}\left(\frac{x-\lambda_j}{\theta_j}\right)\right)} \leqslant \frac{h^{-1}(u_i)F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)}{1-h^{-1}(u_i)\left(1-F^{\alpha_i}\left(\frac{x-\lambda_i}{\theta_i}\right)\right)}.$$

Hence substituting the above results, we get $\frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_i} - \frac{\partial F_{Y_{n:n}}(x)}{\partial \alpha_j} \leq 0$. Thus by Lemma 3.3 of Kundu *et al.* (2016) [24], it can be proved that $F_{Y_{n:n}}(x)$ is Schur-convex in $\underset{\sim}{\alpha}$. thus the result is proved by Theorem A.8 of Marshall *et al.* (2011) [27]. □

Theorem 3.6. *Let Assumption 3.1 hold and $h : [0,1] \to R$ be a differentiable function. Also, $\lambda_i = \lambda_i^* = \lambda$, $\theta_i = \theta_i^* = \theta$, $\alpha_i^* = \beta_i$, where $i = 1, ..., n$. Then,*

(**i**) $\underset{\sim}{\alpha} \overset{w}{\geqslant} \underset{\sim}{\beta}$ *implies* $Y_{1:n} \leqslant_{st} Y_{1:n}^*$, *provided* $\underset{\sim}{\alpha}, \underset{\sim}{\beta} \in \varepsilon_+$, $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$;

(**ii**) $\underset{\sim}{\alpha}^{-1} \geqslant_w \underset{\sim}{\beta}^{-1}$ *implies* $Y_{1:n} \leqslant_{st} Y_{1:n}^*$, *provided* $\underset{\sim}{\alpha}, \underset{\sim}{\beta} \in \varepsilon_+$, $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$.

Proof:

(**i**): The survival function of $Y_{1:n}$ is given by

$$\bar{F}_{Y_{1:n}}(x) = \prod_{i=1}^{n} p_i \left[1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right],$$

where $h(p_i) = u_i$. To prove the required result, it sufficient to show that $\bar{F}_{Y_{1:n}}(x) \leq \bar{F}_{Y_{1:n}^*}(x)$, which is equivalent to proving that $\prod_{i=1}^{n}\left[1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right] \leq \prod_{i=1}^{n}\left[1 - F^{\beta_i}\left(\frac{x-\lambda}{\theta}\right)\right]$, since $\prod_{i=1}^{n} p_i \leq$

$\prod_{i=1}^{n} p_i^*$. Let $\phi(\underset{\sim}{\alpha}) = \prod_{i=1}^{n}\left[1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right]$. Differentiating $\phi(\underset{\sim}{\alpha})$, partially, with respect to α_i, we get

$$\frac{\partial \phi(\underset{\sim}{\alpha})}{\partial \alpha_i} = -\frac{F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)}\phi(\underset{\sim}{\alpha}) \geqslant 0,$$

so, $\phi(\underset{\sim}{\alpha})$ is increasing in each α_i. Again, it can be shown that

$$\begin{aligned}\frac{\partial \phi(\underset{\sim}{\alpha})}{\partial \alpha_i} - \frac{\partial \phi(\underset{\sim}{\alpha})}{\partial \alpha_j} &\overset{\text{sign}}{=} \frac{F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)} \\ &\quad - \frac{F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)}\end{aligned} \tag{3.5}$$

Now,

$$\frac{\partial}{\partial \alpha}\left(\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1 - F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}\right) \overset{\text{sign}}{=} F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right] \leqslant 0,$$

implying that $\frac{F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}{1-F^{\alpha}\left(\frac{x-\lambda}{\theta}\right)}$ is decreasing in α. Thus, as $\underset{\sim}{\alpha} \in \varepsilon_+$, for $i \leqslant j$ taking $\alpha_i \leqslant \alpha_j$, it can be written that

$$\frac{F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)}{1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)} \geqslant \frac{F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)}{1 - F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)},$$

which implies

$$\frac{F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)} \leqslant \frac{F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)\ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{\alpha_j}\left(\frac{x-\lambda}{\theta}\right)}.$$

Hence, from (3.5), we get $\frac{\partial \phi(\underset{\sim}{\alpha})}{\partial \alpha_i} - \frac{\partial \phi(\underset{\sim}{\alpha})}{\partial \alpha_j} \geqslant 0$. Thus, by Lemma 3.3 of Kundu *et al.* (2016) [24], it can be proved that $\phi(\underset{\sim}{\alpha})$ is Schur-concave in $\underset{\sim}{\alpha}$. Thus the result is proved by Theorem A.8 of Marshall *et al.* (2011) [27].

(**ii**): The survival function of $Y_{1:n}$ can be expressed as the function of $c_i = \frac{1}{\alpha_i}$, $i = 1, ..., n$. We denote it by $\psi_5(\underset{\sim}{c})$, where $\underset{\sim}{c} = (c_1, ..., c_n)$, and

$$\psi_5(\underset{\sim}{\alpha}) = \prod_{i=1}^{n} p_i\left[1 - F^{\frac{1}{c_i}}\left(\frac{x-\lambda}{\theta}\right)\right],$$

where $h(p_i) = u_i$. To prove the required result, it is sufficient to show that $\psi_5(\underset{\sim}{\alpha}) \leq \psi_5(\underset{\sim}{\beta})$, which is equivalent to proving that $\prod_{i=1}^{n}\left[1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right] \leq \prod_{i=1}^{n}\left[1 - F^{\beta_i}\left(\frac{x-\lambda}{\theta}\right)\right]$, since $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$. Let $\phi(\underset{\sim}{c}) = \prod_{i=1}^{n}\left[1 - F^{\frac{1}{c_i}}\left(\frac{x-\lambda}{\theta}\right)\right]$. It can be shown that the partial derivative of $\phi(\underset{\sim}{c})$, with respect to c_i, is decreasing in each c_i. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], it can be proved that $\phi(\underset{\sim}{c})$ is Schur-concave in $\underset{\sim}{c}$. Thus the result is proved by Theorem A.8 of Marshall *et al.* (2011) [27]. □

Theorem 3.7. *Let Assumption 3.1 hold and $h : [0, 1] \to R$ be a differentiable function. Also, $\lambda_i = \lambda_i^* = \lambda$, $\theta_i = \theta_i^* = \theta$, and $\alpha_i^* = \beta_i$, where $i = 1, ..., n$. Then, $\underset{\sim}{\alpha} \overset{p}{\geqslant} \underset{\sim}{\beta}$ implies $Y_{1:n} \leqslant_{st} Y_{1:n}^*$, provided $\underset{\sim}{\alpha}, \underset{\sim}{\beta} \in D_+$, $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$.*

Proof: The survival function of $Y_{1:n}$ can be expressed as the function of d_i, where $d_i = \ln \alpha_i$, $i = 1, ..., n$. We denote it by $\psi_6(\underset{\sim}{d})$, where $\underset{\sim}{d} = (d_1, ..., d_n)$:

$$\psi_6(\underset{\sim}{d}) = \prod_{i=1}^{n} p_i \left[1 - F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)\right],$$

where $h(p_i) = u_i$. To prove the required result, it is sufficient to show that $\psi_6(\underset{\sim}{\alpha}) \leq \psi_6(\underset{\sim}{\beta})$, which is equivalent to proving that $\prod_{i=1}^{n} p_i \left[1 - F^{\alpha_i}\left(\frac{x-\lambda}{\theta}\right)\right] \leq \prod_{i=1}^{n} p_i \left[1 - F^{\beta_i}\left(\frac{x-\lambda}{\theta}\right)\right]$, since $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$. Let $\phi(\underset{\sim}{d}) = \prod_{i=1}^{n} \left[1 - F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)\right]$ Differentiating $\psi_6(\underset{\sim}{d})$, partially, with respect to d_i, we get

$$\frac{\partial \phi(\underset{\sim}{d})}{\partial d_i} = -\frac{e^{d_i} F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right) \ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)} \phi(\underset{\sim}{d}) \geqslant 0$$

so, $\phi(\underset{\sim}{d})$ is increasing in each d_i. Again, it can be shown that

$$\frac{\partial \phi(\underset{\sim}{d})}{\partial d_i} - \frac{\partial \phi(\underset{\sim}{d})}{\partial d_j} \overset{\text{sign}}{=} \frac{e^{d_j} F^{e^{d_j}}\left(\frac{x-\lambda}{\theta}\right) \ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{e^{d_j}}\left(\frac{x-\lambda}{\theta}\right)} - \frac{e^{d_i} F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right) \ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)}. \tag{3.6}$$

Now,

$$\frac{\partial}{\partial d}\left(\frac{F^{e^d}\left(\frac{x-\lambda}{\theta}\right)}{1 - F^{e^d}\left(\frac{x-\lambda}{\theta}\right)}\right) \overset{\text{sign}}{=} e^d F^{e^d}\left(\frac{x-\lambda}{\theta}\right) \ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right] \leqslant 0,$$

implying that $\frac{F^{e^d}\left(\frac{x-\lambda}{\theta}\right)}{1-F^{e^d}\left(\frac{x-\lambda}{\theta}\right)}$ is decreasing in d. Thus, as $\underset{\sim}{d} \in D_+$, for $i \leqslant j$ taking $d_i \geqslant d_j$, it can be written that

$$\frac{F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)}{1 - F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)} \leqslant \frac{F^{e^{d_j}}\left(\frac{x-\lambda}{\theta}\right)}{1 - F^{e^{d_j}}\left(\frac{x-\lambda}{\theta}\right)},$$

which implies

$$\frac{e^{d_i} F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right) \ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{e^{d_i}}\left(\frac{x-\lambda}{\theta}\right)} \geqslant \frac{e^{d_j} F^{e^{d_j}}\left(\frac{x-\lambda}{\theta}\right) \ln\left[F\left(\frac{x-\lambda}{\theta}\right)\right]}{1 - F^{e^{d_j}}\left(\frac{x-\lambda}{\theta}\right)}.$$

Hence, from (3.6), we get $\frac{\partial \phi(\underset{\sim}{d})}{\partial d_i} - \frac{\partial \phi(\underset{\sim}{d})}{\partial d_j} \leqslant 0$. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], it can be proved that $\phi(\underset{\sim}{d})$ is Schur-concave in $\underset{\sim}{d}$. Thus the result is proved by Lemma 3.1 of Khaledi *et al.* (2002) [16]. □

Theorem 3.8. *Let Assumption 3.1 hold and $h : [0, 1] \to R$ be a differentiable function. Also, $\lambda_i = \lambda_i^* = \lambda$, $\theta_i^* = \delta_i$, and $\alpha_i = \alpha_i^* = \alpha$, where $i = 1, ..., n$. Then, $\underset{\sim}{\theta}^{-1} \geqslant_w \underset{\sim}{\delta}^{-1}$ implies $Y_{1:n} \leqslant_{st} Y_{1:n}^*$, provided $\underset{\sim}{\theta}, \underset{\sim}{\delta} \in \varepsilon_+$, $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$, and $\tilde{r}(u)$ is increasing in u.*

Proof: The survival function of $Y_{1:n}$ can be expressed as the function of e_i, where $e_i = \frac{1}{\theta_i}$, $i = 1, ..., n$. We denote it by $\psi_7(\underset{\sim}{e})$, where $\underset{\sim}{e} = (e_1, ..., e_n)$, and

$$\psi_7(\underset{\sim}{e}) = \prod_{i=1}^{n} p_i \left[1 - F^{\alpha}\big(e_i(x-\lambda)\big)\right],$$

where $h(p_i) = u_i$. To prove the required result, it is sufficient to show that $\psi_7(\underset{\sim}{\theta}^{-1}) \leq \psi_7(\underset{\sim}{\delta}^{-1})$, which is equivalent to proving that $\prod_{i=1}^{n}\left[1 - F^{\alpha}\left(\frac{1}{\theta_i}(x-\lambda)\right)\right] \leq \prod_{i=1}^{n}\left[1 - F^{\alpha}\left(\frac{1}{\delta_i}(x-\lambda)\right)\right]$, since $\prod_{i=1}^{n} p_i \leq \prod_{i=1}^{n} p_i^*$. Let $\phi(\underset{\sim}{e}) = \prod_{i=1}^{n}\left[1 - F^{\alpha}\left(e_i(x-\lambda)\right)\right]$. Differentiating $\phi(\underset{\sim}{e})$, partially, with respect to e_i, we get

$$\frac{\partial \phi(\underset{\sim}{e})}{\partial e_i} = -\frac{\alpha(x-\lambda)\tilde{r}(e_i(x-\lambda))F^{\alpha}(e_i(x-\lambda))}{1 - F^{\alpha}(e_i(x-\lambda))}\phi(\underset{\sim}{e}) \leqslant 0,$$

so, $\psi_7(\underset{\sim}{e})$ is decreasing in each e_i. Again, it can be shown that

$$\begin{aligned}\frac{\partial \phi(\underset{\sim}{e})}{\partial e_i} - \frac{\partial \phi(\underset{\sim}{e})}{\partial e_j} &\overset{\text{sign}}{=} \frac{\alpha(x-\lambda)\tilde{r}(e_j(x-\lambda))F^{\alpha}(e_j(x-\lambda))}{1 - F^{\alpha}(e_j(x-\lambda))} \\ &\quad - \frac{\alpha(x-\lambda)\tilde{r}(e_i(x-\lambda))F^{\alpha}(e_i(x-\lambda))}{1 - F^{\alpha}(e_i(x-\lambda))}.\end{aligned} \tag{3.7}$$

Now,

$$\frac{\partial}{\partial e}\left(\frac{F^{\alpha}(e(x-\lambda))}{1 - F^{\alpha}(e(x-\lambda))}\right) \overset{\text{sign}}{=} \alpha(x-\lambda)\tilde{r}(e(x-\lambda))F^{\alpha}(e(x-\lambda)) \geqslant 0,$$

implying that $\frac{F^{\alpha}(e(x-\lambda))}{1-F^{\alpha}(e(x-\lambda))}$ is inscreasing in e. Thus, us $\underset{\sim}{\theta} \in \varepsilon_+$, for $i \leqslant j$ taking $e_i \geqslant e_j$, it can be written that

$$\frac{F^{\alpha}(e_i(x-\lambda))}{1 - F^{\alpha}(e_i(x-\lambda))} \geqslant \frac{F^{\alpha}(e_j(x-\lambda))}{1 - F^{\alpha}(e_j(x-\lambda))}.$$

As $\tilde{r}(u)$ is increasing in u, then

$$\frac{\alpha(x-\lambda)\tilde{r}(e_i(x-\lambda))F^{\alpha}(e_i(x-\lambda))}{1 - F^{\alpha}(e_i(x-\lambda))} \geqslant \frac{\alpha(x-\lambda)\tilde{r}(e_j(x-\lambda))F^{\alpha}(e_j(x-\lambda))}{1 - F^{\alpha}(e_j(x-\lambda))}.$$

Hence, from (3.7), we get $\frac{\partial \phi(\underset{\sim}{e})}{\partial e_i} - \frac{\partial \phi(\underset{\sim}{e})}{\partial e_j} \leqslant 0$. Thus, by Lemma 3.1 of Kundu *et al.* (2016) [24], it can be proved that $\phi(\underset{\sim}{e})$ is Schur-concave in $\underset{\sim}{e}$. Thus the result is proved by Theorem A.8 Marshall *et al.* (2011) [27]. □

Theorem 3.8 shows that if both the location and shock parameter vectors i.e. the matrix of location and shock parameters of one system majorizes the other when the scale and shape parameter vectors remain constant do not lead to better system reliability.

Theorem 3.9. *For $n = 2$, let Assumption 3.1 hold. Further, let $h : [0,1] \to R_+$ be a differentiable and strictly increasing concave function. Then, for $i = 1, 2$, if $\theta_i = \theta_i^* = \alpha_i = \alpha_i^* = \theta$, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) \in P_2$, we have that*

$$\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \end{pmatrix},$$

implies $Y_{1:2}^* \geqslant_{st} Y_{1:2}$, provided $\tilde{r}(u)$ is increasing in u.

Proof: With $u_1 = h(p_1)$, $u_2 = h(p_2)$, we have $p_1 = h^{-1}(u_1), p_2 = h^{-1}(u_2)$, where h^{-1} denotes the inverse of the function h. From (1.2), the survival function of $Y_{1:2}$ is

$$\bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta) = \prod_{i=1}^{2} h^{-1}(u_i)\left[1 - F^{\theta}\left(\frac{x-\lambda_i}{\theta}\right)\right], \qquad x > \max\{\lambda_i,\, i = 1, ..., n\}.$$

Note that the function $\bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)$ is permutation invariant in (u_i, λ_i), and so condition (i) of Theorem 2 of Balakrishnan *et al.* (2015) [1] is satisfied. Next, we have to show that condition (ii) of Theorem 2 of Balakrishnan *et al.* (2015) [1] also holds. The assumption $(\underset{\sim}{u},\underset{\sim}{\lambda}) \in P_2$ implies that $(u_1 - u_2)(\lambda_1 - \lambda_2) \leqslant 0$. This implies that $u_1 \geqslant u_2$ and $\lambda_1 \leqslant \lambda_2$ or $u_1 \leqslant u_2$ and $\lambda_1 \geqslant \lambda_2$. We proof only for the case when $u_1 \geqslant u_2$ and $\lambda_1 \leqslant \lambda_2$. The proof for the other case is similar. The partial derivatives of $\bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)$ with respect to u_i and λ_i are

$$\begin{aligned}
\frac{\partial \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)}{\partial u_i} &= \frac{\frac{\partial h^{-1}(u_i)}{\partial u_i}}{h^{-1}(u_i)} \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta),\\
\frac{\partial \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)}{\partial \lambda_i} &= -\frac{\tilde{r}\left(\frac{x-\lambda_i}{\theta}\right) F^{\theta}\left(\frac{x-\lambda_i}{\theta}\right)}{1 - F^{\theta}\left(\frac{x-\lambda_i}{\theta}\right)} \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta).
\end{aligned}$$

For fixed $x > \max\{\lambda_i,\, i = 1, ..., n\}$, let us define the function $\varphi(\underset{\sim}{u},\underset{\sim}{\lambda})$ as follows:

$$\begin{aligned}
\varphi(\underset{\sim}{u},\underset{\sim}{\lambda}) &= (u_1 - u_2)\left(\frac{\partial \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)}{\partial u_1} - \frac{\partial \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)}{\partial u_2}\right)\\
&\quad + (\lambda_1 - \lambda_2)\left(\frac{\partial \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)}{\partial \lambda_1} - \frac{\partial \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)}{\partial \lambda_2}\right)\\
&= (u_1 - u_2)\left(\frac{\frac{\partial h^{-1}(u_1)}{\partial u_1}}{h^{-1}(u_1)} - \frac{\frac{\partial h^{-1}(u_2)}{\partial u_2}}{h^{-1}(u_2)}\right) \times \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta)\\
&\quad + (\lambda_1 - \lambda_2)\left(\frac{\tilde{r}\left(\frac{x-\lambda_1}{\theta}\right) F^{\theta}\left(\frac{x-\lambda_1}{\theta}\right)}{1 - F^{\theta}\left(\frac{x-\lambda_1}{\theta}\right)} - \frac{\tilde{r}\left(\frac{x-\lambda_2}{\theta}\right) F^{\theta}\left(\frac{x-\lambda_2}{\theta}\right)}{1 - F^{\theta}\left(\frac{x-\lambda_2}{\theta}\right)}\right) \bar{F}_{Y_{1:2}}(x;\underset{\sim}{u},\underset{\sim}{\lambda},\theta,\theta). \qquad (3.8)
\end{aligned}$$

Since h is strictly increasing and concave, then for $u_1 \geqslant u_2$ and $\lambda_1 \leqslant \lambda_2$, we have

$$\frac{\frac{\partial h^{-1}(u_1)}{\partial u_1}}{h^{-1}(u_1)} \leq \frac{\frac{\partial h^{-1}(u_2)}{\partial u_2}}{h^{-1}(u_2)}. \qquad (3.9)$$

Furthermore, $\frac{t^{\theta}}{1-t^{\theta}}$ is increasing in t for $\theta > 0$. For the reversed hazard rate function $\tilde{r}$ that is increasing, we have

$$\tilde{r}(\frac{x-\lambda_1}{\theta}) \geq \tilde{r}\left(\frac{x-\lambda_2}{\theta}\right) \qquad (3.10)$$

and

$$\frac{\tilde{r}\left(\frac{x-\lambda_1}{\theta}\right) F^{\theta}\left(\frac{x-\lambda_1}{\theta}\right)}{1 - F^{\theta}\left(\frac{x-\lambda_1}{\theta}\right)} \geqslant \frac{\tilde{r}\left(\frac{x-\lambda_2}{\theta}\right) F^{\theta}\left(\frac{x-\lambda_2}{\theta}\right)}{1 - F^{\theta}\left(\frac{x-\lambda_2}{\theta}\right)}. \qquad (3.11)$$

Combining (3.9) and (3.10) and (3.11), we see that $\varphi(\underset{\sim}{u},\underset{\sim}{\lambda}) \leqslant 0$. Condition (ii) Theorem 2 of Balakrishnan *et al.* (2015) [1] is satisfied, which completes the proof. □

Counterexample 3.1. Let the baseline distribution function be $F(t) = e^{\frac{-1}{t}}$, $t > 0$. Take $h(p) = -\ln p$. Here, the baseline reversed hazard rate function is decreasing and $h(p)$ is decreasing and convex. Thus, the assumptions of the Theorem 3.9 are violated. Let us take $\theta_1 = \theta_2 = \theta_1^* = \theta_2^* = \alpha_1 = \alpha_2 = \alpha_1^* = \alpha_2^* = 1.5$, $(\lambda_1, \lambda_2) = (0.3, 0.9)$, $(\lambda_1^*, \lambda_2^*) = (0.54, 0.66)$, $(p_1, p_2) = (e^{-0.4}, e^{-0.5})$ and $(p_1^*, p_2^*) = (e^{-0.44}, e^{-0.46})$. Consider the T-transform matrix $T_{0.6} = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$. It can be shown that

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \end{pmatrix} T_{0.6},$$

which implies $\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \end{pmatrix}$. Under this set up, $\bar{F}_{Y_{1:2}}(2) = 0.2597610428$, $\bar{F}_{Y_{1:2}^*}(2) = 0.2599036428$, $\bar{F}_{Y_{1:2}}(5) = 0.06532417018$, $\bar{F}_{Y_{1:2}^*}(5) = 0.06516286182$, which readily shows that $Y_{1:2}^* \not\geqslant_{st} Y_{1:2}$.

The following theorem extends Theorem 3.8 when two sets of n-independent observations from ELS distribution. The generalization is the direct result of the Theorem 3.8 and Lemma 5 of Balakrishnan *et al.* (2018) [2]. So, the proof is omitted.

Theorem 3.10. *Let Assumption 3.1 hold and $h : [0, 1] \to R_+$ be a differentiable and strictly increasing concave function. Further, let T_w be a T-transform matrix. Then, for $i = 1, ..., n$, if $\theta_i = \theta_i^* = \alpha_i = \alpha_i^* = \theta$, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) \in P_n$, we have that*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix} T_w,$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$, provided $\tilde{r}(u)$ is increasing in u.

Theorem 3.11. *Let Assumption 3.1 hold. Further, let $T_{w_1}, ..., T_{w_k}$ have the same structure. Suppose $h : [0, 1] \to R_+$ is a differentiable and strictly increasing concave function. Then, for $i = 1, ..., n$, if $\theta_i = \theta_i^* = \alpha_i = \alpha_i^* = \theta$, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) \in P_n$, we have that*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix} T_{w_1} \cdots T_{w_k},$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$, provided $\tilde{r}(u)$ is increasing in u.

Proof: Since a finite product of T-transform matrices with the same structure is also a T-transform matrix, so, the desired result is obtained from Theorem 3.9. □

Our next Theorem shows that the result in Theorem 3.10 holds for T-transform matrices with different structure.

Theorem 3.12. *Let Assumption 3.1 hold. Further, let $T_{w_1}, ..., T_{w_k}, k > 2$, have different structures. Suppose $h : [0, 1] \to R_+$ is a differentiable and strictly increasing concave function. Then, for $i = 1, ..., n$, if $\theta_i = \theta_i^* = \alpha_i = \alpha_i^* = \theta$, $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) \in P_n$ and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) T_{w_1} \cdots T_{w_k} \in P_n$, we have that*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \end{pmatrix} T_{w_1} \cdots T_{w_k},$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$, provided $\tilde{r}(u)$ is increasing in u.

Proof:

$$\begin{pmatrix} h(p_1)^{(i)} & \cdots & h(p_n)^{(i)} \\ \lambda_1^{(i)} & \cdots & \lambda_n^{(i)} \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \end{pmatrix} T_{w_1} \cdots T_{w_i}, \qquad \text{for} \quad i = 1, ..., k.$$

Assume $V_1^{(i)}, ..., V_n^{(i)}$, $i = 1, ..., k$, are independent sets of random variables with $V_j^{(i)} \sim ELS(\lambda_j^{(i)}, \theta_j, \alpha_j)$ where $\theta_j = \theta_j^* = \alpha_j = \alpha_j^* = \theta$, $j = 1, ..., n$ and $i = 1, ..., k$. From the assumption of the theorem, it follows that

$$\begin{pmatrix} h(p_1)^{(i)} & \cdots & h(p_n)^{(i)} \\ \lambda_1^{(i)} & \cdots & \lambda_n^{(i)} \end{pmatrix} \in P_n, \qquad \text{for} \quad i = 1, ..., k.$$

From these observations and the results of Theorem 3.9, it then follows that

$$Y_{1:n} \leqslant_{st} V_{1:n}^{(1)} \leqslant_{st} \cdots \leqslant_{st} V_{1:n}^{(k-2)} \leqslant_{st} V_{1:n}^{(k-1)} \leqslant_{st} Y_{1:n}^*,$$

which completes the proof of the theorem. □

The following example illustrates the result established in Theorem 3.11.

Example 3.1. Suppose X_1, X_2 and X_3 are independent non-negative random variables with $X_i \sim ELS(\lambda_i, \theta_i, \alpha_i)$, and I_{p_1}, I_{p_2} and I_{p_3} are independent Bernoulli random variables, independent of X_i's, with $E(I_{p_i}) = p_i, i = 1, 2, 3$. Further, suppose X_1^*, X_2^* and X_3^* are independent non-negative random variables with $X_i^* \sim ELS(\lambda_i^*, \theta_i^*, \alpha_i^*)$, and $I_{p_1^*}, I_{p_2^*}$ and $I_{p_3^*}$ are independent Bernoulli random variables, independent of $X_i^{*'}s$, with $E(I_{p_i^*}) = p_i^*$, $i = 1, 2, 3$. Consider a baseline distribution with distribution function $F(x) = 1 - e^{-x}$, $x > 0$. Consider the T-transform matrices as follows:

$$T_{0.7} = \begin{pmatrix} 0.7 & 0 & 0.3 \\ 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \end{pmatrix}, \quad T_{0.6} = \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad T_{0.4} = \begin{pmatrix} 0.6 & 0 & 0.4 \\ 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \end{pmatrix}.$$

Suppose $h(p) = \frac{p}{1+p}$. Then, for $\theta_i = \theta_i^* = \alpha_i = \alpha_i^* = 1$, $i = 1, 2, 3$, let $(\lambda_1, \lambda_2, \lambda_3) = (2, 3, 4)$, $(\lambda_1^*, \lambda_2^*, \lambda_3^*) = (3.06, 2.76, 3.17)$, $(p_1, p_2, p_3) = (0.4, 0.3, 0.2)$ and $(p_1^*, p_2^*, p_3^*) = (0.285, 0.317, 0.273)$. It is easy to observe that $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}) \in P_3, (h(\underset{\sim}{p}), \underset{\sim}{\lambda})T_{0.7} \in P_3$ and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda})T_{0.7}T_{0.6} \in P_3$ and $(h(\underset{\sim}{p^*}), \underset{\sim}{\lambda^*}) = (h(\underset{\sim}{p}), \underset{\sim}{\lambda})T_{0.7}T_{0.6}T_{0.4}$. So, from Theorem 3.11, we have $Y_{1:3}^* \geqslant_{st} Y_{1:3}$.

Theorem 3.13. *Let Assumption 3.1 hold for $n = 2$. Suppose $h : [0, 1] \to R_+$ is differentiable and strictly increasing concave function. Further, let $\tilde{r}(u)$ and $u\tilde{r}(u)$ increasing in u. Then, if $\alpha_i = \alpha_i^* = \alpha$, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \in S_2$, we have that*

$$\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \\ \theta_1^* & \theta_2^* \end{pmatrix}$$

implies $Y_{1:2}^* \geqslant_{st} Y_{1:2}$.

Proof: With $u_1 = h(p_1), u_2 = h(p_2)$, we have $p_1 = h^{-1}(u_1)$, $p_2 = h^{-1}(u_2)$, where h^{-1} denotes the inverse of the function h. From (1.2), the survival function of $Y_{1:2}$ is

$$\bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha) = \prod_{i=1}^{2} h^{-1}(u_i)\left[1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)\right], \qquad x > \max\{\lambda_i,\, i = 1, ..., n\}.$$

Note that the function $\bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)$ is permutation invariant in $(u_i, \lambda_i, \theta_i)$ and therefore condition (i) of Lemma 6 of Balakrishnan *et al.* (2018) [2] is satisfied. Next, we have to show that Condition (ii) of Lemma 6 of Balakrishnan *et al.* (2018) [2] also holds. The assumption $(\underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \in S_2$ implies that $u_1 \leqslant (\geqslant) u_2$ and $\lambda_1 \geqslant (\leqslant)\lambda_2$ and $\theta_1 \geqslant (\leqslant)\theta_2$. We proof only for the case when $u_1 \leqslant u_2$ and $\lambda_1 \geqslant \lambda_2$ and $\theta_1 \geqslant \theta_2$. The proof for the other case is similar. The partial derivatives of $\bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)$ with respect to u_i and λ_i and θ_i are

$$\begin{aligned}
\frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial u_i} &= \frac{\frac{\partial h^{-1}(u_i)}{\partial u_i}}{h^{-1}(u_i)} \times \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha), \\
\frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial \lambda_i} &= \frac{\alpha}{\theta_i} \times \frac{\tilde{r}\left(\frac{x-\lambda_i}{\theta_i}\right) F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)}{1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)} \times \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha), \\
\frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial \theta_i} &= \frac{\alpha}{\theta_i} \times \frac{\left(\frac{x-\lambda_i}{\theta_i}\right)\tilde{r}\left(\frac{x-\lambda_i}{\theta_i}\right) F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)}{1 - F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)} \times \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha).
\end{aligned}$$

For fixed $x > \max\{\lambda_i,\, i = 1, ..., n\}$, let us define the function $\varphi(\underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta})$ as follows:

$$\begin{aligned}
\varphi(\underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}) &= (u_1 - u_2)\left(\frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial u_1} - \frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial u_2}\right) \\
&\quad + (\lambda_1 - \lambda_2)\left(\frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial \lambda_1} - \frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial \lambda_2}\right) \\
&\quad + (\theta_1 - \theta_2)\left(\frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial \theta_1} - \frac{\partial \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha)}{\partial \theta_2}\right) \\
&= (u_1 - u_2)\left(\frac{\frac{\partial h^{-1}(u_1)}{\partial u_1}}{h^{-1}(u_1)} - \frac{\frac{\partial h^{-1}(u_2)}{\partial u_2}}{h^{-1}(u_2)}\right) \times \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha) \\
&\quad + (\lambda_1 - \lambda_2)\left(\frac{1}{\theta_1}\tilde{r}\left(\frac{x-\lambda_1}{\theta_1}\right) l_1 - \frac{1}{\theta_2}\tilde{r}\left(\frac{x-\lambda_2}{\theta_2}\right) l_2\right) \times \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha) \\
&\quad + (\theta_1 - \theta_2)\left(\frac{1}{\theta_1}\left(\frac{x-\lambda_1}{\theta_1}\right)\tilde{r}\left(\frac{x-\lambda_1}{\theta_1}\right) l_1 - \frac{1}{\theta_2}\left(\frac{x-\lambda_2}{\theta_2}\right)\tilde{r}\left(\frac{x-\lambda_2}{\theta_2}\right) l_2\right) \\
&\quad \times \bar{F}_{Y_{1:2}}(x; \underset{\sim}{u}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \alpha),
\end{aligned} \tag{3.12}$$

where $l_i = l\left(\alpha, F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)\right) = \frac{\alpha F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)}{1-F^{\alpha}\left(\frac{x-\lambda_i}{\theta_i}\right)}$, $i = 1, 2$, is defined in Lemma 2.8 of Torrado (2015) [28]. Since h is strictly increasing and concave function, then for $u_1 \leqslant u_2$ and $\lambda_1 \geqslant \lambda_2$ and $\theta_1 \geqslant \theta_2$, we have

$$\frac{\frac{\partial h^{-1}(u_1)}{\partial u_1}}{h^{-1}(u_1)} \geqslant \frac{\frac{\partial h^{-1}(u_2)}{\partial u_2}}{h^{-1}(u_2)}. \tag{3.13}$$

$$\frac{1}{\theta_1}\tilde{r}\left(\frac{x-\lambda_1}{\theta_1}\right)l\left(\alpha, F^{\alpha}\left(\frac{x-\lambda_1}{\theta_1}\right)\right) \leqslant \frac{1}{\theta_2}\tilde{r}\left(\frac{x-\lambda_2}{\theta_2}\right)l\left(\alpha, F^{\alpha}\left(\frac{x-\lambda_2}{\theta_2}\right)\right) \tag{3.14}$$

and

$$\frac{1}{\theta_1}\left(\frac{x-\lambda_1}{\theta_1}\right)\tilde{r}\left(\frac{x-\lambda_1}{\theta_1}\right)l\left(\alpha, F^{\alpha}\left(\frac{x-\lambda_1}{\theta_1}\right)\right) \leqslant \frac{1}{\theta_2}\left(\frac{x-\lambda_2}{\theta_2}\right)\tilde{r}\left(\frac{x-\lambda_2}{\theta_2}\right)l\left(\alpha, F^{\alpha}\left(\frac{x-\lambda_2}{\theta_2}\right)\right) \tag{3.15}$$

combining (3.13), (3.14) and (3.15) in (3.12), we see that $\varphi(\underset{\sim}{y}, \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \leqslant 0$. So condition (ii) of Lemma 6 of Balakrishnan *et al.* (2018) [2] is satisfied, which completes the proof. □

Counterexample 3.2. Let the baseline distribution function be $F(t) = e^{\frac{-1}{t}}$, $t > 0$. Here, $\tilde{r}(t)$ and $t\tilde{r}(t)$ are decreasing. Take $h(p) = p$. Thus, the assumption of Theorem 3.13 are violated. Let us take $\alpha_1 = \alpha_2 = \alpha_1^* = \alpha_2^* = 2.2$, $(\theta_1, \theta_2) = (5.2, 2.7)$, $(\theta_1^*, \theta_2^*) = (4.2, 3.7)$, $(\lambda_1, \lambda_2) = (2.2, 2.5)$, $(\lambda_1^*, \lambda_2^*) = (2.32, 2.38)$, $(p_1, p_2) = (p_1^*, p_2^*) = (0.2, 0.2)$. Consider the T-transform matrix $T_{0.6} = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$. It can be shown that

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \\ \theta_1^* & \theta_2^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \end{pmatrix} T_{0.6},$$

which implies $\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \\ \theta_1^* & \theta_2^* \end{pmatrix}$. Under this set up, $\bar{F}_{Y_{1:2}}(2.8) = 0.9999999923$, $\bar{F}_{Y_{1:2}^*}(2.8) = 0.9999999918$, $\bar{F}_{Y_{1:2}}(5) = 0.8918294672$, $\bar{F}_{Y_{1:2}^*} = 0.9248659543$, which readily shows that $Y_{1:2}^* \not\geqslant_{st} Y_{1:2}$.

The following theorem extends Theorem 3.12 when two sets of n-independent observations are from ELS distribution. The generalization is the direct of the Theorem 3.12 and Lemma 7 of Balakrishnan *et al.* (2018) [2]. So, the proof is omitted.

Theorem 3.14. *Let Assumption 3.1 hold. Suppose $h : [0, 1] \to R_+$ is differentiable and strictly increasing concave function. Further, let $\tilde{r}(u)$ and $u\tilde{r}(u)$ are increasing in u. Then, for $i = 1, ..., n$ and T-transform matrix T_w, if $\alpha_i = \alpha_i^* = \alpha$, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \in S_n$, we have that*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \\ \theta_1^* & \theta_2^* & \cdots & \theta_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \theta_1 & \theta_2 & \cdots & \theta_n \end{pmatrix} T_w$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$.

Theorem 3.15. *Let Assumption 3.1 hold and $T_{w_1}, ..., T_{w_k}$ be T-transform matrices with same structures. Suppose, $h : [0, 1] \to R_+$ is differentiable and strictly increasing concave*

function. Further, let $\tilde{r}(u)$, *and* $u\tilde{r}(u)$ *are strictly increasing in* u,. *Then, if* $\alpha_i = \alpha_i^* = \alpha$,, *and* $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \in S_n$, *we have that*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \\ \theta_1^* & \theta_2^* & \cdots & \theta_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \theta_1 & \theta_2 & \cdots & \theta_n \end{pmatrix} T_{w_1} \cdots T_{w_k}$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$.

Proof: Since, a finite product of T-transform matrices with the same structure is also a T-transform matrix, so, the desired result can be obtained by repeating the result of Theorem 3.13. □

Theorem 3.16. *Let Assumption 3.1 hold and* $T_{w_1}, ..., T_{w_k}$, $k > 2$ *be T-transform matrices with different structures. Suppose,* $h : [0, 1] \to R_+$ *is differentiable and strictly increasing concave function. Further, let* $\tilde{r}(u)$ *and* $u\tilde{r}(u)$ *are increasing in* u. *Then, if* $\alpha_i = \alpha_i^* = \alpha$, *and* $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \in S_n$ *and* $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta})\, T_{w_1} \cdots T_{w_i} \in S_n, i = 1, ..., k-1$, *we have*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \\ \theta_1^* & \theta_2^* & \cdots & \theta_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \theta_1 & \theta_2 & \cdots & \theta_n \end{pmatrix} T_{w_1} \cdots T_{w_k}$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$.

Proof:

$$\begin{pmatrix} h^{(i)}(p_1) & \dots & h^{(i)}(p_n) \\ \lambda_1^{(i)} & \cdots & \lambda_n^{(i)} \\ \theta_1^{(i)} & \cdots & \theta_n^{(i)} \end{pmatrix} = \begin{pmatrix} h(p_1) & \cdots & h(p_n) \\ \lambda_1 & \cdots & \lambda_n \\ \theta_1 & \cdots & \theta_n \end{pmatrix} T_{w_1} ... T_{w_i}, \qquad \text{for} \quad i = 1, ..., k.$$

Assume $V_1^{(i)}, ..., V_n^{(i)}$, $i = 1, ..., k$, are independent sets of random variables with $V_j^{(i)} \sim ELS(\lambda_j^{(i)}, \theta_j^{(i)}, \alpha_j)$ where $\alpha_i = \alpha_i^* = \alpha$, $j = 1, ..., n$ and $i = 1, ..., k$. From the assumption of the theorem, it follows that

$$\begin{pmatrix} h(p_1)^{(i)} & \dots & h(p_n)^{(i)} \\ \lambda_1^{(i)} & \cdots & \lambda_n^{(i)} \\ \theta_1^{(i)} & \cdots & \theta_n^{(i)} \end{pmatrix} \in S_n.$$

Using the results of Theorem 3.13, it then follows that

$$Y_{1:n} \leqslant_{st} V_{1:n}^{(1)} \leqslant_{st} \cdots \leqslant_{st} V_{1:n}^{(k-2)} \leqslant_{st} V_{1:n}^{(k-1)} \leqslant_{st} Y_{1:n}^*,$$

which completes the proof of the theorem. □

Theorem 3.17. *Let Assumption 3.1 hold for $n = 2$. Suppose $h : [0, 1] \to R_+$ is a differentiable and strictly increasing concave function. Further, assume that $\tilde{r}(u)$ and $u\tilde{r}(u)$ are increasing in u, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha}) \in U_2$,. Then,*

$$\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \\ \theta_1^* & \theta_2^* \\ \alpha_1^* & \alpha_2^* \end{pmatrix}$$

implies $Y_{1:2}^* \geqslant_{st} Y_{1:2}$.

Proof: With the help of Lemma 2.1, the proof follows from arguments similar to those in the proof of Theorem 3.13. It is omitted for brevity. □

We present a counterexample to show that the comparison result may not hold if the assumptions are not satisfied.

Counterexample 3.3. Let the baseline distribution function be $F(t) = 1 - \exp(1 - t^{0.5})$, $t \geq 1$. Here $\tilde{r}(t)$ and $t\tilde{r}(t)$ are decreasing. Take $h(p) = e^p$, where $h(p)$ is convex. Thus, the assumption of Theorem 3.17 are not violated. Let us set $(\alpha_1, \alpha_2) = (0.2, 0.5)$, $(\alpha_1^*, \alpha_2^*) = (0.44, 0.26)$, $(\lambda_1, \lambda_2) = (1, 1.5)$, $(\lambda_1^*, \lambda_2^*) = (1.4, 1.1)$, $(\theta_1, \theta_2) = (4, 2)$, $(\theta_1^*, \theta_2^*) = (2.4, 3.6)$, $(p_1, p_2) = (\ln(4), \ln(5))$, $(p_1^*, p_2^*) = (\ln(4.8), \ln(4.2))$,. Consider a T-transform matrix $T_{0.2} = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}$. Then, it can be shown that

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \\ \theta_1^* & \theta_2^* \\ \alpha_1^* & \alpha_2^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} T_{0.2},$$

which implies $\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \lambda_1^* & \lambda_2^* \\ \theta_1^* & \theta_2^* \\ \alpha_1^* & \alpha_2^* \end{pmatrix}$. Finally $\bar{F}_{Y_{1:2}}(5) - \bar{F}_{Y_{1:2}^*}(5) = 0.4133022299$, $\bar{F}_{Y_{1:2}}(7.5) - \bar{F}_{Y_{1:2}^*}(7.5) = -0.0324688838$, which readily shows that $Y_{1:2}^* \not\geqslant_{st} Y_{1:2}$.

In the following theorem, we present a generalization of Theorem 3.17 to the case of n independent variables.

Theorem 3.18. *Let Assumption 3.1 hold. Further, let T_w be a T-transform matrix. Suppose $h : [0, 1] \to R_+$ is a differentiable strictly increasing concave function. Further, assume that $\tilde{r}(u)$ and $u\tilde{r}(u)$ are increasing in u, and let $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha}) \in U_n$. Then,*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \\ \theta_1^* & \theta_2^* & \cdots & \theta_n^* \\ \alpha_1^* & \alpha_2^* & \cdots & \alpha_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \theta_1 & \theta_2 & \cdots & \theta_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix} T_w$$

implies $Y_{1:n}^* \geqslant_{st} Y_{1:n}$.

Theorem 3.19. *Let $T_{w_1}, ..., T_{w_k}$ be T-transform matrices with same structure. Let Assumption 3.1 hold and $h : [0,1] \to R_+$ be a differentiable and strictly increasing concave function. Further, assume that $\tilde{r}(u)$ and $u\tilde{r}(u)$ are strictly increasing in u, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha}) \in U_n$. Then,*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \\ \theta_1^* & \theta_2^* & \cdots & \theta_n^* \\ \alpha_1^* & \alpha_2^* & \cdots & \alpha_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \theta_1 & \theta_2 & \cdots & \theta_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix} T_{w_1} \cdots T_{w_k}$$

implies $Y^*_{1:n} \geqslant_{st} Y_{1:n}$.

The following theorem presents a generalization to the case of a finite number of T-transform matrices with different structures.

Theorem 3.20. *Let Assumption 3.1 hold. Further, let $T_{w_1}, ..., T_{w_k}$, $k > 2$ be T-transform matrices, with different structures. Suppose $h : [0,1] \to R_+$ is a differentiable and strictly increasing concave function. Further, assume that $\tilde{r}(u)$ and $u\tilde{r}(u)$ are increasing in u, and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha}) \in U_n$ and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}, \underset{\sim}{\alpha})\, T_{w_1} \cdots T_{w_i} \in U_n, i = 1, ..., k-1$. Then,*

$$\begin{pmatrix} h(p_1^*) & h(p_2^*) & \cdots & h(p_n^*) \\ \lambda_1^* & \lambda_2^* & \cdots & \lambda_n^* \\ \theta_1^* & \theta_2^* & \cdots & \theta_n^* \\ \alpha_1^* & \alpha_2^* & \cdots & \alpha_n^* \end{pmatrix} = \begin{pmatrix} h(p_1) & h(p_2) & \cdots & h(p_n) \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \theta_1 & \theta_2 & \cdots & \theta_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix} T_{w_1} \cdots T_{w_k}$$

implies $Y^*_{1:n} \geqslant_{st} Y_{1:n}$.

4. SOME SPECIAL CASES

In this section, we present some special cases of the results obtained in the previous sections. We consider two special distributions - generalized gamma, half - normal distributions. For these distributions, we present some comparisons results using the general results established earlier. In terms of hazard rate function, the following theorems can be proved like the theorems above, which are proved by reversed hazard rate function. To prevent recurrence, you can refer to Das *et al.* (2021) [7].

4.1. Generalized gamma distribution

In this subsection, we consider the generalized gamma distribution with density function

$$f(t) \propto t^{a-1} e^{-t^b}, \qquad a, b, t > 0.$$

It is easy to check that the hazard rate function of this distribution is increasing for $a, b \geq 1$ and decreasing for $0 < a, b \leq 1$(see Hazra *et al.*, 2018 [15]). We now consider the baseline distribution function to the generalized gamma distribution.

Theorem 4.1. *For a baseline distribution function $F(.)$, let $X_1, ..., X_n$ $(X_1^*, ..., X_n^*)$ be non-negative independent random variables with $X_i \sim ELS(\lambda_i, \theta_i, \alpha_i)[X_i^* \sim ELS(\mu_i, \delta_i, \beta_i)]$, $i = 1, ..., n$. Further, let $I_{p_1}, ..., I_{p_n}[I_{p_1^*}, ..., I_{p_n^*}]$ be a set of independent Bernoulli random variables, independent of $X_i[X_i^*]$'s with $E(I_{p_i}) = p_i[E(I_{p_i^*}) = p_i^*]$, $i = 1, ..., n$. Further, let $h : [0, 1] \to R_+$ be a differentiable, increasing and convex function. Then, for $i = 1, 2$, if $\theta_i = \delta_i = \delta$ and $\alpha_i = \beta_i = \alpha \geq 1$, $r(x)$ is increasing. Suppose $h(p) = e^p \ln(1 + p)$. Then, for $a, b \geq 1$, we have $\begin{pmatrix} h(p_1) & h(p_2) \\ \theta_1 & \theta_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \delta_1 & \delta_2 \end{pmatrix}$ implies $Y_{1:2}^* \geq_{st} Y_{1:2}$.*

Theorem 4.2. *For a baseline distribution function $F(.)$, let $X_1, ..., X_n$ $(X_1^*, ..., X_n^*)$ be non-negative independent random variables with $X_i \sim ELS(\lambda_i, \theta_i, \alpha_i)[X_i^* \sim ELS(\mu_i, \delta_i, \beta_i)]$, $i = 1, ..., n$. Further, let $I_{p_1}, ..., I_{p_n}[I_{p_1^*}, ..., I_{p_n^*}]$ be a set of independent Bernoulli random variables, independent of $X_i[X_i^*]$'s with $E(I_{p_i}) = p_i[E(I_{p_i^*}) = p_i^*]$, $i = 1, ..., n$. Further, let $h : [0, 1] \to R_+$ be a differentiable, increasing and convex function. Further, let the baseline hazard rate $r(.)$ be increasing. If $\alpha_i = \beta_i = \alpha \geq 1$ and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\theta}) \in N_2$. Suppose $h(p) = p^2$. Then, for $a, b \geq 1$, we have $\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \theta_1 & \theta_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \mu_1 & \mu_2 \\ \delta_1 & \delta_2 \end{pmatrix}$ implies $Y_{1:2}^* \geq_{st} Y_{1:2}$.*

4.2. Half-normal distribution

Consider the probability distribution function of a half-normal distribution given by

$$f(t) \propto e^{\frac{-t^2}{2}}, \qquad t > 0.$$

The hazard rate function of the above half-normal distribution is increasing (see Hazra *et al.* (2018) [15]). The distribution function of the half-normal distribution is now taken as the baseline distribution function.

Theorem 4.3. *For a baseline distribution function $F(.)$, let $X_1, ..., X_n$ $(X_1^*, ..., X_n^*)$ be non-negative independent random variables with $X_i \sim ELS(\lambda_i, \theta_i, \alpha_i)$ $(X_i^* \sim ELS(\mu_i, \delta_i, \beta_i))$, $i = 1, ..., n$. Further, let $I_{p_1}, ..., I_{p_n}[I_{p_1^*}, ..., I_{p_n^*}]$ be a set of independent Bernoulli random variables, independent of $X_i[X_i^*]$'s with $E(I_{p_i}) = p_i[E(I_{p_i^*}) = p_i^*]$, $i = 1, ..., n$. Further, let $h : [0, 1] \to R_+$ be a differentiable, increasing and convex function. Further, let the baseline hazard rate $r(.)$ be increasing. If $\theta_i = \delta_i = \theta$ and $(h(\underset{\sim}{p}), \underset{\sim}{\lambda}, \underset{\sim}{\alpha}) \in N_2^*$. Let $h(p) = -p \ln(1 - p)$. Then $\begin{pmatrix} h(p_1) & h(p_2) \\ \lambda_1 & \lambda_2 \\ \alpha_1 & \alpha_2 \end{pmatrix} \gg \begin{pmatrix} h(p_1^*) & h(p_2^*) \\ \mu_1 & \mu_2 \\ \beta_1 & \beta_2 \end{pmatrix}$ implies $Y_{1:2}^* \geq_{st} Y_{1:2}$.*

5. APPLICATIONS

In this section, we discuss application of few of our established results in insurance and auction theory. Suppose $X_1, ..., X_n$ are independent exponentiated location-scale random variables with $X_i \sim ELS(\lambda_i, \mu_i, \alpha_i)$, for $i = 1, ..., n$, and $I_{p_1}, ..., I_{p_n}$ are independent Bernoulli random variables, independent of the X_i's, with $E(I_{p_i}) = p_i$. Let $Y_i = I_{p_i} X_i$, for $i = 1, ..., n$.

Suppose X_i denotes the total of random claims that can be made in an insurance period and I_{p_i} denotes a Bernoulli random variables associated with X_i defined as follows: $I_{p_i} = 1$ when ever the ith policyholder makes random claim X_i and $I_{p_i} = 0$ whenever he/she does not make make a claim. In setting, $Y_i = I_{p_i} X_i$ corresponds to the claim amount in a portfolio of risks. The problem of comparison of number of claims and aggregate claim amounts with respect to some well-known stochastic orders is of interest on both theoretical and practical view points. Under some conditions Theorems 3.2, 3.3, 3.4 respectively conclude that $Y_{n:n}$ in the weakly supermajorized order, weakly submajorized order, p-larger and reciprocally majorized order is stochastically smaller. There are many real-life applications of the ordering results. We discuss applications of few of our established results in auction theory. Auction theory has been an interest topic to various scientists because of its usefulness for sale of variety of items or purchasing services. For more details in auction theory, we refer to (Klemperer, (2004) [17]). In real world,among all types of auctions, the sealed-bid private-value auction is of theoretical interest. Also, this type of auction has been used extensively. In this case, bidders hand in their bids to the auctioneer simultaneously and can neither observe their rival bids nor revise their own bids.The bidders having the highest bid wins. The bidders with the lowest bid wins in the reverse auction. Consequently, the bidder pays his own bid in the sealed-bid first-price auction (FPA). Few of our established results could be useful for some new light in the auction theory. Let the bids follow exponentiated locatio-scale model. Then, under some conditions, Theorems 3.2, 3.3, 3.4 respectively conclude that the final price in the FPA with more heterogeneous shape parameters in the weakly super majorized order, reciprocal of the shape parameters in the weakly submajorized, scale parameters in the p-larger and reciprocally majorized orders is stochastically smaller.

6. CONCLUDING

In this paper, when the matrix of parameters changes to another matrix of parameters with respect to multivariate chain majorization, we study the usual stochastic order of the smallest order statistics when each component receives a random shock. Under certain conditions, by using the concept of vector majorization and related orders, we have also discussed stochastic comparison between series and parallel systems in the sense of the usual stochastic order under random shock. We have then applied the results for some special cases of the exponentiated location-scale model with possibly different scale, location and shape parameters to illustrate the established results.

ACKNOWLEDGMENTS

The authors are thankful to the Editor and an anonymous referee for some useful comments and suggestions, which have resulted in improving the content and the presentation of the paper.

REFERENCES

[1] Balakrishnan, N.; Haidari, A. and Masoumifard, K. (2015). Stochastic comparisons of series and parallel systems with generalized exponential components, *IEEE Transaction on Reliability*, **64**(1), 333–348.

[2] Balakrishnan, N.; Nanda, P. and Kayal, S. (2018). Ordering of series and parallel systems comprising heterogeneous generalized modified Weibull components, *Applied Stochastic Models in Business and Industry*, **34**(6), 816–834.

[3] Barmalzan, G.; Najafabadi, A.T.P. and Balakrishnan, N. (2017). Ordering properties of the smallest and largest claim amounts in a general scale model, *Scandinavian Actuarial Journal*, **2017**(2), 105–124.

[4] Chowdhury, S. and Kundu, A. (2017). Stochastic comparisons of parallel systems with log-Lindley distributed components, *Operations Research Letters*, **45**(3), 199–205.

[5] Chowdhury, S. and Kundu, A. (2018). *Stochastic comparisons of parallel systems with log-Lindley distributed components under random shocks*, arXiv preprint, arXiv: 1804.04103.

[6] Chowdhury, S.; Kundu, A. and Mishra, S.K. (2022). On comparison of two parallel systems having Log-Lindley distributed components, *Communication in Statistics – Theory and Methods*, **52**(1), 121–140.

[7] Das, S.; Kayal, S. and Chowdhury, D. (2021). Ordering results on extremes of exponentiated location-scale models, *Probability in the Engineering and Informational Sciences*, **35**(2), 331–354.

[8] Das, S. and Kayal, S. (2020). Ordering extremes of exponentiated location-scale models with dependent and heterogeneous random samples, *Metrika*, **83**, 869–893.

[9] Dykstra, R.; Kochar, SC. and Rojo, J. (1997). Stochastic comparisons of parallel systems of heterogeneous exponential components, *Journal of Statistical Planning and Inference*, **65**, 203–211.

[10] Fang, L. and Balakrishnan, N. (2018). Ordering properties of the smallest order statistics from generalized Birnbaum–Saunders models with associated random shocks, *Metrika*, **81**(1), 19–35.

[11] Fang, R.; Li, C. and Li, X. (2016). Stochastic comparisons on sample extremes of dependent and heterogeneous observations, *Statistics*, **50**(4), 930–955.

[12] Fang, L. and Zhang, X. (2013). Stochastic comparisons of series systems with heterogeneous Weibull components, *Statistic and Probability Letters*, **83**(7), 1649–1653.

[13] Fang, L. and Xu, T. (2019). Ordering results of the smallest and largest order statistics from independent heterogeneous exponentiated gamma random variables, *Statistica Neerlandica*, **73**(2), 197–210.

[14] Hazra, N.K.; Kuiti, M.R.; Finkelstein, M. and Nanda, A.K. (2017). On stochastic comprarisons of maxima order statistics from the location-scale family of distributions, *Journal of multivariate Analysis*, **16**, 31–41.

[15] Hazra, N.K.; Kuiti, M.R.; Finkelstein, M. and Nanda, A.K. (2018). On stochastic comparisons of minimum order statistics from the location-scale family of distributions, *Metrika*, **81**(2), 105–123.

[16] Khaledi, B.E. and Kochar, S.C. (2002). Dispersive ordering among linear combination of uniform random variables, *Journal of Statistical planning and Inference*, **100**(1), 13–21.

[17] Klemperer, P. (2004). *Auction: Theory and Practice*, New York, Academic Press.

[18] Kochar, S.C. and Torrado, N. (2015). On stochastic comparisons of largest order statistics in the scale model, *Communications in Statistics – Theory and Methods*, **44**(19), 4132–4143.

[19] Kundu, A. and Chowdhury, S. (2016). Ordering properties of order statistics from heterogeneous exponentiated Weibull models, *Statistics and Probability Letters*, **114**, 119–127.

[20] Kundu, A. and Chowdhury, S. (2018). Ordering properties of sample minimum from Kumaraswamy-G random variables, *Statistics*, **52**(1), 133–146.

[21] Kundu, A. and Chowdhury, S. (2019). Ordering properties of the largest order statistics from Kumaraswamy-G models under random shocks, *Communication in Statistics – Theory and Method*, **50**(6), 1502–1514.

[22] Kundu, A. and Chowdhury, S. (2020). On stochastic comparisons of series systems with heterogeneous dependent and independent location-scale family distributed components, *Operation Research Letters*, **48**(1), 40–47.

[23] Kundu, A.; Chowdhury, S. and Balakrishnan, N. (2023). Ordering properties of the smallest and largest lifetimes in Gompertz-Makeham model, *Communication in Statistics – Theory and Methods*, **52**(3), 643–669.

[24] Kundu, A.; Chowdhury, S.; Nanda, A.K. and Hazra, N.K. (2016). Some results on majorization and their applications, *Journal of Computational and Applied Mathematics*, **301**, 161–177.

[25] Li, C. and Li, X. (2015). Likelihood ratio order of sample minimum from heterogeneous Weibull random variables, *Statistics and Probability Letters*, **97**, 46–53.

[26] Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, **84**, 641–652.

[27] Marshall, A.W.; Olkin, I. and Arnold, B.C. (2011). *Inequalities: Theory of Majorization and Its Applications*, Springer Series in Statistics, New York.

[28] Torrado, N. (2015). On magnitude ordering between smallest order statistics from heterogeneous beta distributions, *Journal of Mathematical Analysis and Applications*, **426**(2), 824–838.

[29] Torrado, N. and Kochar, S.C. (2015). Stochastic order relations among parallel systems from Weibull distributions, *Journal of Applied Probability*, **52**(1), 102–116.

[30] Zhao, P. and Balakrishnan, N. (2011). Some characterization results for parallel systems with two heterogeneous exponential components, *Statistics*, **45**(6), 593–604.

REVSTAT-Statistical journal

Aims and Scope

The aim of REVSTAT-Statistical Journal is to publish articles of high scientific content, developing Statistical Science focused on innovative theory, methods, and applications in different areas of knowledge. Important survey/review contributing to Probability and Statistics advancement is also welcome.

Background

Statistics Portugal started in 1996 the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, a quarterly publication whose goal was the publication of papers containing original research results, and application studies, namely in the economic, social and demographic fields. Statistics Portugal was aware of how vital statistical culture is in understanding most phenomena in the present-day world, and of its responsibilities in disseminating statistical knowledge.

In 1998 it was decided to publish papers in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work. At the time, the editorial board was mainly composed by Portuguese university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal. In 2001, the *Revista de Estatística* published a three volumes special issue containing extended abstracts of the invited and contributed papers presented at the 23rd European Meeting of Statisticians (EMS). During the EMS 2001, its editor-in-chief invited several international participants to join the editorial staff.

In 2003 the name changed to REVSTAT-Statistical Journal, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

Editorial policy

REVSTAT-Statistical Journal is an open access peer-reviewed journal published quarterly, in English, by Statistics Portugal.

The editorial policy of REVSTAT is mainly placed on the originality and importance of the research. The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage revstat.ine.pt based in Open Journal System (OJS). The only working language allowed is English. Authors intending to submit any work must register, login and follow the guidelines.

There are no fees for publishing accepted manuscripts that will be made available in open access.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external. Authors can suggest an editor or reviewer who is expert on the paper subject providing her/his complete information, namely: name, affiliation, email and, if possible, personal URL or ORCID number.

All published works are Open Access (CC BY 4.0) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Also, in the context of archiving policy, REVSTAT is a *blue* journal welcoming authors to deposit their works in other scientific repositories regarding the use of the published edition and providing its source.

Journal prints may be ordered at expenses of the author(s), and prior to publication.

Abstract and Indexing services

REVSTAT-Statistical Journal is covered *by Journal Citation Reports - JCR (Clarivate); DOAJ-Directory Of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; Scimago Journal & Country Rank; Scopus*

Author guidelines

The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage https://revstat.ine.pt/ based in Open Journal System (OJS). Authors intending to submit any work must *register, login* and follow the indications choosing *Submissions*.

REVSTAT - Statistical Journal adopts the COPE guidelines on publication ethics.

Work presentation

- the only working language is English;
- the first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s);
- a summary of fewer than one hundred words, followed by a maximum of six keywords and the MSC 2020 subject classification should be included also in the first page;
- manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and a maximum of 25 pages;
- the title should be with no more than 120 characters (with spaces);
- figures must be a minimum of 300dpi and will be reproduced online as in the original work, however, authors should take into account that the printed version is always in black and grey tones;
- authors are encouraged to submit articles using LaTeX which macros are available at *REVSTAT style*;
- citations in text should be included in the text by name and year in parentheses, as in the following examples: § article title in lowercase (Author 1980); § This

theorem was proved later by AuthorB and AuthorC (1990); § This subject has been widely addressed (AuthorA 1990; AuthorB et al. 1995; AuthorA and AuthorB 1998).

- references should be listed in alphabetical order of the author's scientific surname at the end of the article;
- acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text;
- authors are welcome to suggest one of the Editors or Associate Editors or yet other reviewer expert on the subject providing a complete information, namely: name, affiliation, email and personal URL or ORCID number in the Comments for the Editor (submission form).

Accepted papers

After final revision and acceptance of an article for publication, authors are requested to provide the corresponding LaTeX file, as in REVSTAT style.

Supplementary files may be included and submitted separately in .tiff, .gif, .jpg, .png .eps, .ps or .pdf format. These supplementary files may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

Copyright Notice

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information.

According to REVSTAT's *archiving policy*, after assigning the copyright form, authors may cite and use limited excerpts (figures, tables, etc.) of their works accepted/published in REVSTAT in other publications and may deposit only the published edition in scientific repositories providing its source as REVSTAT while the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.